

Riccardo Brumm

Energieprofilbasierende Stabilitätsanalyse
von Membranproteinen

BACHELORARBEIT

HOCHSCHULE MITTWEIDA

UNIVERSITY OF APPLIED SCIENCES

Mathematik/Naturwissenschaften/Informatik

Mittweida, 2010

Riccardo Brumm

Energieprofilbasierende Stabilitätsanalyse
von Membranproteinen

eingereicht als

BACHELORARBEIT

an der

HOCHSCHULE MITTWEIDA

UNIVERSITY OF APPLIED SCIENCES

Mathematik/Naturwissenschaften/Informatik

Mittweida, 2010

Erstprüfer:	Prof. Dr. rer. Nat. Dirk Labudde
Zweitprüfer:	Dipl. Inf. (FH) Daniel Stockmann

Bibliografische Beschreibung

Brumm, Riccardo:

Energieprofilbasierte Stabilitätsanalyse von Membranproteinen. – 2010. – 65 S.
Mittweida, Hochschule Mittweida, Fakultät Mathematik/Naturwissenschaften/Informatik,
Bachelorarbeit, 2010

Referat:

Es ist eines der größten ungelösten Probleme der letzten Jahrzehnte, wie die Beziehung zwischen Sequenz und 3D-Struktur in Proteinen tatsächlich aussieht. Um die Funktion und den Einfluss von physiologischen Bedingungen exakt verstehen zu können, werden Aussagen über die Struktur von Proteinen benötigt. Energieprofile stellen ein Energiemodell dar, welches die damit verbundenen Aspekte teilweise untersucht. Das Hauptaugenmerk liegt dabei auf Wechselwirkungen der einzelnen Aminosäuren eines Proteins untereinander. Die Grundlage bildet die Berechnung der freien Energie jeder Aminosäure. Zu Beginn dieser Arbeit war es nur möglich, von globulären Proteinen Energieprofile zu berechnen.

Ziel der Bachelorarbeit ist es, einen Algorithmus zu entwickeln, der von Membranproteinen Energieprofile berechnen kann, um verstehen zu können, wie sich unterschiedliche physiologische Bedingungen auf Membranproteine auswirken, und wie beispielsweise Punktmutationen die Funktionsfähigkeit und den strukturellen Aufbau von Membranproteinen beeinflussen können.

Inhaltsverzeichnis

Abbildungsverzeichnis	ii
Tabellenverzeichnis	iii
1 Einleitung	1
1.1 Motivation.....	2
2 Grundlagen	3
2.1 Proteine und ihre Strukturen	3
2.1.1 Die Kanonischen Aminosäuren und ihre Eigenschaften	3
2.1.2 Primärstruktur, Sekundärstruktur, Tertiärstruktur.....	6
2.2 Interatomare Kräfte und Proteinstruktur	10
2.2.1 Kovalente Wechselwirkungen.....	10
2.2.2 Wasserstoffbrückenbindungen.....	11
2.2.3 Hydrophobe und hydrophile Wechselwirkungen.....	11
2.2.4 Ladungs-Ladungs-, Ladungs-Dipol- u. Dipol-Dipol-Wechselwirkungen.....	12
2.2.5 Van-der-Waals-Kräfte.....	13
2.2.6 Abstoßende Kräfte.....	13
2.3 Ein Einblick in die Proteinfaltung – Die Energielandschaft.....	14
3 Energiefunktion bei Membranproteinen	18
3.1 Wechselwirkungen mit der Umgebung.....	18
3.2 Wechselwirkungen innerhalb des Proteins.....	21
3.3 Ein Paarpotential zur Beschreibung wechselwirkender Aminosäuren.....	22
3.4 Ergänzende Information zur Energiefunktion.....	23
4 Evaluierung der Energiefunktion	24
4.1 Erstellung meines Datensatzes.....	24
4.2 Validierung des Berechnungsalgorithmus	26
4.3 Präferenzen zu Aminosäuren von Membranproteinen im Datensatz.....	28
4.3.1 Vergleich Präferenzen globulärer Proteine und Membranproteine.....	30
4.3.2 Vergleich Präferenzen innerhalb und außerhalb der Membran.....	31
4.4 Vergleich des Berechnungsalgorithmus mit experimentellen Daten	32
5 Ergebnisse und Analyse der Energiefunktion	33
5.1 Allgemeine energetische Analyse von Membranproteinen.....	33
5.2 Energetische Analyse von Membran- und Nichtmembranbereichen	35
5.3 Energetische Analyse der Sekundärstrukturelemente	38
5.4 Energetische Charakterisierung von Aminosäuren	40
5.4.1 Anlegen energetischer Präferenzen von Aminosäuren	40
5.4.2 Auswertung der energetischen Präferenzen.....	42
5.4.3 Korrelationsmatrizen der Aminosäuren.....	43
6 Ausblick	46
Anlagen	47
Literaturverzeichnis	52
Danksagung	54
Selbständigkeitserklärung	55

Abbildungsverzeichnis

Abbildung 1: Venn-Diagramm der Aminosäuren.....	4
Abbildung 2: Schematische Darstellung eines Rückgrats eines Proteins.....	6
Abbildung 3: Darstellung der α -Helix.	8
Abbildung 4: Ein schematisch dargestelltes Protein mit den Sekundärstrukturelementen.....	9
Abbildung 5: Molekülgefüge von Chloratomen.	10
Abbildung 6: Molekülmodell Glycerin mit Elektronenwolken.....	11
Abbildung 7: Darstellung sterischer Wechselwirkungen zwischen Atomen.	13
Abbildung 8: Erste Vorstellung des Faltungstrichters nach Levinthal.	14
Abbildung 9: Schematische Darstellung des Faltungstrichters eines Proteins.....	15
Abbildung 10: Grundformen der Energielandschaften.	16
Abbildung 11: Energieprofile in einem Energie-Sequenz-Diagramm.....	17
Abbildung 12: Die „POPS area per atom“ Ausgabe eines Proteins in einer Datei gespeichert.	19
Abbildung 13: XML-Auszug eines Proteins aus der PDBTM	24
Abbildung 14: Vergleich von 1bj0 in der PDBTM und der PDBsum.....	25
Abbildung 15: Vergleich beider Innen/Außen-Kriterien an einem Protein.	28
Abbildung 16: Energieprofil von Bacteriorhodopsin (PDB-ID: 1brr).....	32
Abbildung 17: Energetische Verteilung aller Aminosäuren im Datensatz.	33
Abbildung 18: E- Δ E-SASA Plot aller Aminosäuren.....	34
Abbildung 19: Energetische Verteilung äußerer bzw. innerer Aminosäuren.....	35
Abbildung 20: E- Δ E-SASA Plot aller Aminosäuren, die innerhalb der Membran liegen.....	36
Abbildung 21: E- Δ E-SASA Plot aller Aminosäuren, die außerhalb der Membran liegen.	37
Abbildung 22: Energieprofil eines Proteins mit der PDB-ID 1a0s.....	38
Abbildung 23: Vergleich der Energieverteilung von Asparaginsäure und Glutaminsäure.	44
Abbildung 24: Vergleich der Energieverteilung von Glycin und Lysin.....	45

Tabellenverzeichnis

Tabelle 1: Die relevanten Eigenschaften der Aminosäuren im Überblick.....	5
Tabelle 2: Längenwerte der einzelnen Aminosäuren.....	19
Tabelle 3: Häufigkeiten der Zustände der einzelnen Aminosäuren für Membranproteine.	21
Tabelle 4: Häufigkeiten der einzelnen Aminosäuren im Datensatz.	22
Tabelle 5: Vergleich des ursprünglichen mit dem neu entwickelten Innen/Außen-Kriterium.....	26
Tabelle 6: Vergleich beider Innen/Außen-Kriterien an einem Protein.....	27
Tabelle 7: Verteilung der Aminosäuren im Datensatz.	29
Tabelle 8: Präferenzen der kanonischen Aminosäuren in den Sekundärstrukturelementen.	30
Tabelle 9: Statistik zur energetischen Verteilung der Sekundärstruktur.....	39
Tabelle 10: Statistik zur energetischen Verteilung der Sekundärstruktur relativ zur Membran.	39
Tabelle 11: Präferenzen der Energiewerte von Aminosäuren innerhalb der Membran.....	40
Tabelle 12: Präferenzen der Energiewerte von Aminosäuren außerhalb der Membran.	41
Tabelle 13: Gesamte Statistik der Energiewerte der Aminosäuren.....	47
Tabelle 14: Korrelationsmatrix für innerhalb der Membran stehender Aminosäuren.	51
Tabelle 15: Korrelationsmatrix für außerhalb der Membran stehender Aminosäuren.....	51

1 Einleitung

Ohne Membranen wäre Leben in der uns bekannten Form nicht möglich. Sie sind selektiv permeabel für Ionen und impermeabel für Makromoleküle. Mit Membranen ist es Zellen möglich, von der Umgebung abweichende Bedingungen in ihrem Inneren zu schaffen. Zusätzlich können einzelne Bereiche innerhalb der Zellen voneinander abgetrennt werden. In diesen sogenannten Zellkompartimenten können wiederum andere Bedingungen herrschen. Ein Beispiel dafür sind Mitochondrien, in denen der Hauptteil des von der Zelle genutzten ATPs produziert wird. Diese unterschiedlichen physiologischen Bedingungen zwischen einzelnen Zellkompartimenten bzw. dem Zellinneren und dem extrazellulären Raum, sind durch gesteuerten aktiven und passiven Transport von Ionen durch die Membran möglich. Dieser Transport ist durch integrale Membranproteine gewährleistet, welche die gesamte Membran durchdringen. Ein Beispiel hierfür ist der Transport von Natrium- und Kaliumionen durch die Membran. Die resultierenden Konzentrationsunterschiede zwischen den Zellen und ihrer Umgebung sind für die Zellen essentiell. Somit haben Membranproteine eine enorm große Bedeutung für die physiologischen Vorgänge in den Zellen.

Ca. 30 % des menschlichen Genoms verschlüsselt Membranproteine [1]. Oft reichen schon Punktmutationen, um die Stabilität und/oder die Funktion eines Proteins einzuschränken. Dabei kann es zu schwerwiegenden Beeinträchtigungen des Organismus kommen. So verursachen einzelne Mutationen einzelner Aminosäuren im menschlichen Rhodopsin, welches sensitiv auf Licht reagiert, *Retinitis pigmentosa* (Degeneration der Netzhaut) [2]. Weitere Mutationsbedingte Krankheiten sind Mukoviszidose [3], Morbus Alzheimer [4] und Krebs [5].

1.1 Motivation

Es ist eines der größten ungelösten Probleme der letzten Jahrzehnte, wie die Beziehung zwischen Sequenz und 3D-Struktur in Proteinen tatsächlich aussieht. Um die Funktion und den Einfluss von physiologischen Bedingungen exakt verstehen zu können, werden Aussagen über die Struktur von Proteinen benötigt. Energieprofile stellen ein Energiemodell dar, welches die damit verbundenen Aspekte teilweise untersucht. Das Hauptaugenmerk liegt dabei auf Wechselwirkungen der einzelnen Aminosäuren eines Proteins untereinander. Die Grundlage bildet die Berechnung der freien Energie jeder Aminosäure.

Die ersten Analysen sowie der Algorithmus zur Berechnung der freien Energie der Aminosäuren in globulären Proteinen, wurden von Dr. Frank Dressel im Zusammenhang mit seiner Dissertationsschrift an der Technischen Universität Dresden entwickelt. Dieser Algorithmus diente als Grundlage und wurde in dieser Arbeit weiterentwickelt, damit auch Energieprofile von Membranproteinen berechnet werden können.

Ziel ist, verstehen zu können, wie sich unterschiedliche physiologische Bedingungen auf Membranproteine auswirken und wie beispielsweise Punktmutationen die Funktionsfähigkeit und den strukturellen Aufbau von Membranproteinen beeinflussen können.

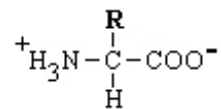
2 Grundlagen

2.1 Proteine und ihre Strukturen

Protein ist eine abgeleitete Sammelbezeichnung für alle natürlich vorkommenden Copolymere, die sich in der Regel aus 20 verschiedenen Aminosäuren als Monomere zusammensetzen. Ab etwa 100 Monomer-Einheiten spricht man allgemein von Proteinen. Diese sind in der belebten Welt, neben Kohlenhydraten und Fetten, die dritte große Gruppe von Nahrungs- und Reservestoffen. Auf der Anwesenheit bestimmter Proteine beruhen Struktur, Funktion und Stoffwechsel aller lebender Zellen und Gewebe. Man findet sie gleichermaßen in Tieren, Pflanzen und Mikroorganismen. [6]

2.1.1 Die Kanonischen Aminosäuren und ihre Eigenschaften

Die 20 am Aufbau von Proteinen beteiligten L- α -Aminosäuren bezeichnet man als kanonische Aminosäuren. Sie sind in reinem Zustand farblose, kristalline Stoffe, die in festem Zustand und in neutraler wässriger Lösung überwiegend als innere Salze (Zwitterionen) vorliegen:



Das bewirkt hohe Schmelztemperaturen und geringe Löslichkeit in unpolaren Lösungsmitteln.

Die Hälfte aller kanonischen Aminosäuren trägt eine unpolare Seitenkette. Diese unpolaren Aminosäuren, dargestellt in Abbildung 1, sind schlecht wasserlöslich und neigen aufgrund des hydrophoben Effekts zur Aggregation.

Glycin als kleinste Aminosäure bekommt eine Sonderstellung, weil „Engpässe“ in Proteinstrukturen oft nur diesen kleinstmöglichen Rest zulassen. Alanin, Leucin, Valin, Isoleucin, Methionin und Prolin besitzen unpolare aliphatische Seitenketten. Alanin, Leucin und Valin besitzen unterschiedlich viele Kohlenstoffatome in ihren Seitenketten, dabei sind Leucin und Isoleucin Konstitutionsisomere. Methionin trägt eine unpolare Thioethergruppe (-S-CH₃) in der Seitenkette und Prolin besitzt als einzige Aminosäure eine sekundäre α -Aminogruppe (-NH-), an der sich die Seitenkette zu einem Pyrrolidin-Heterozyklus schließt. Damit wird die Konformationsfreiheit von Prolin eingeschränkt, wodurch auch die Proteinfaltung limitiert wird. Diese fein abgestuften Größen der hydrophoben Seitenketten

erlauben ein „passgerechtes“ Ausfüllen des Innenraumes von Proteinen. Die Seitenketten von Tryptophan, Phenylalanin und Tyrosin sind von großem Umfang und aromatisch. Phenylalanin ist deutlich hydrophober als Tyrosin und Tryptophan aufgrund der Hydroxylgruppe bzw. des Stickstoffatoms.

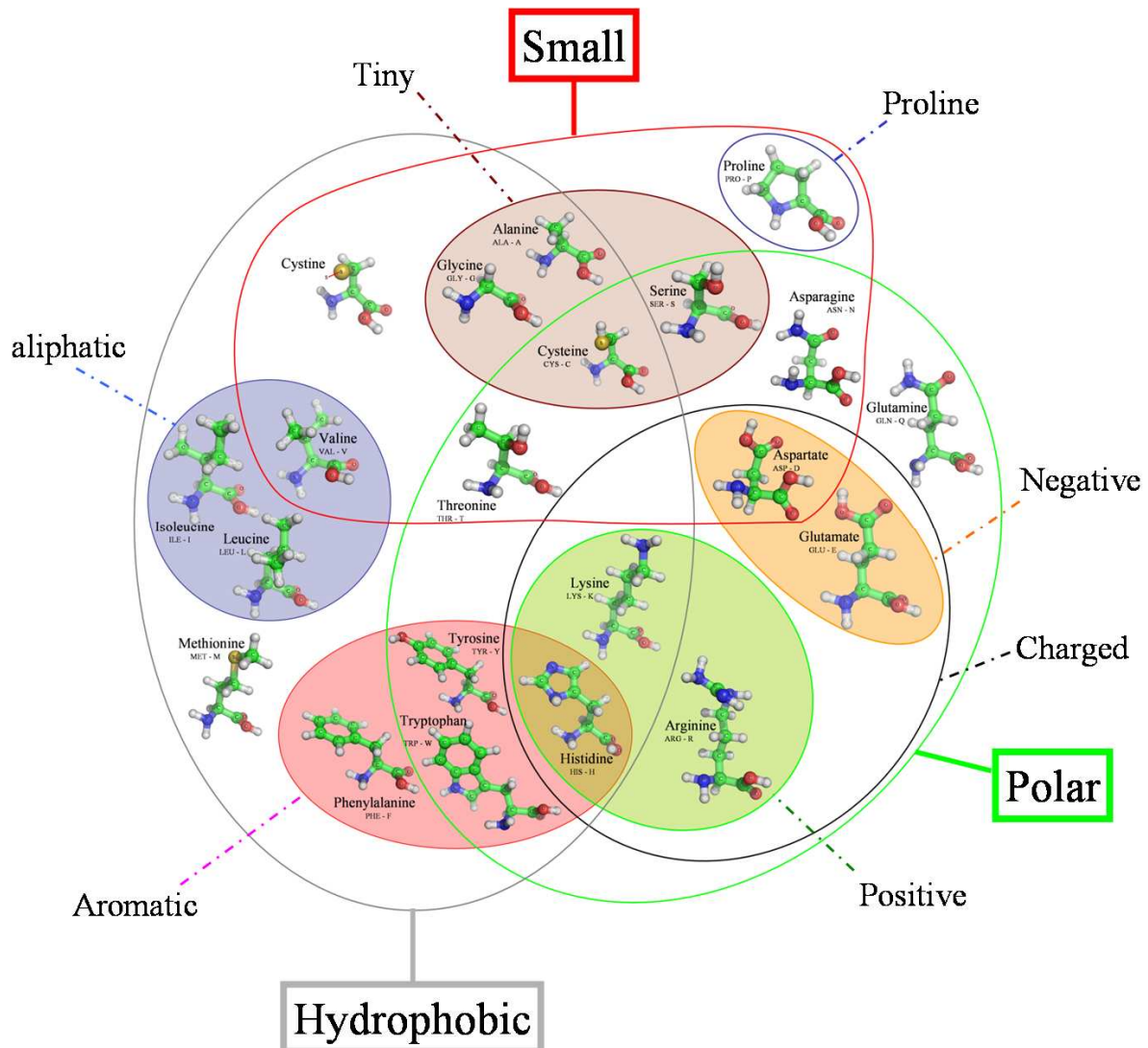


Abbildung 1: Venn-Diagramm der Aminosäuren. [7]

Polare, hydrophobe Seitenketten tragen Serin, Asparagin, Cystein, Threonin und Glutamin. Durch ihre funktionellen Gruppen werden Wasserstoffbrücken zum umgebenden Wasser ausgebildet. Threonin und Serin besitzen je eine Hydroxylgruppe an ihrer Seitenkette. Glutamin und Asparagin tragen an ihrer Seitenkette eine Carboxamidgruppe ($-\text{CONH}_2$). Durch die Thiolgruppe ($-\text{SH}$) bekommt Cystein eine polare Natur. Somit können zwei Cysteine unter Oxidation ihrer Thiole eine kovalente Disulfidbrücke ($-\text{C-S-S-C}-$) ausbilden. Disulfidbrücken ermöglichen kovalente Bindungen zur Stabilisierung innerhalb eines Proteins bzw. auch zwischen Proteinen. Aminosäuren mit geladenen Seitenketten sind im Vergleich zu allen weiteren kanonischen Aminosäuren am besten mit Wasser verträglich

(hydrophil). Asparaginsäure und Glutaminsäure weisen negativ geladene Seitenketten auf, die durch im physiologischen pH-Bereich deprotonierte Carboxylgruppen hervorgerufen werden. Die ionisierten Formen werden Aspartat und Glutamat bezeichnet. Meist positiv geladen sind die Seitenketten basischer Aminosäuren bei physiologischem pH-Wert. Die funktionellen Gruppen in den Seitenketten von Lysin und Arginin sind stark basisch und daher bei neutralem pH-Wert immer protoniert. Histidin weist eine Imidazolgruppe auf und ist zugleich die einzige Aminosäure, die eine Seitenkette trägt, deren pH-Wert nahe dem Neutralpunkt liegt. Sobald die direkte Umgebung innerhalb eines Proteins, in dem sich Histidin befindet, den effektiven pH-Wert verändert, kann der Imidazolrest am Histidin sauer oder basisch werden. Daher dient Histidin in vielen katalytischen Reaktionen als Protonendonator bzw. -akzeptor. [6] Die für die energetische Analyse der Aminosäuren wichtigsten Eigenschaften sind in Tabelle 1 dargestellt.

Tabelle 1: Die relevanten Eigenschaften der Aminosäuren im Überblick. [6], [8], [9], [10]
Zusätzlich zu den Eigenschaften sind 1-Letter-Code, 3-Letter-Code und die Anzahl jeder Aminosäure im Datensatz angegeben.

Aminosäure	3-Letter-Code	1-Letter-Code	Struktur	Hydrophobizität	Ladung	Polarität	Anzahl im Datensatz
Alanin	Ala	A	aliphatisch	hydrophob	ungeladen	unpolar	5286
Arginin	Arg	R	aliphatisch	hydrophil	geladen	polar	2404
Asparagin	Asn	N	aliphatisch	hydrophil	ungeladen	polar	2250
Asparaginsäure	Asp	D	aliphatisch	hydrophil	geladen	polar	2247
Cystein	Cys	C	aliphatisch	hydrophob	ungeladen	polar	706
Glutamin	Gln	Q	aliphatisch	hydrophil	ungeladen	polar	1735
Glutaminsäure	Glu	E	aliphatisch	hydrophil	geladen	polar	2347
Glycin	Gly	G	aliphatisch	hydrophob	ungeladen	unpolar	4770
Histidin	His	H	aromatisch	hydrophob	geladen	polar	1012
Isoleucin	Ile	I	aliphatisch	hydrophob	ungeladen	unpolar	4182
Leucin	Leu	L	aliphatisch	hydrophob	ungeladen	unpolar	6474
Lysin	Lys	K	aliphatisch	hydrophob	geladen	polar	2382
Methionin	Met	M	aliphatisch	hydrophob	ungeladen	unpolar	1513
Phenylalanin	Phe	F	aromatisch	hydrophob	ungeladen	unpolar	3440
Prolin	Pro	P	heterocyclisch	hydrophil	ungeladen	unpolar	2454
Serin	Ser	S	aliphatisch	hydrophil	ungeladen	polar	3642
Threonin	Thr	T	aliphatisch	hydrophob	ungeladen	polar	3343
Tryptophan	Trp	W	aromatisch	hydrophob	ungeladen	unpolar	1157
Tyrosin	Tyr	Y	aromatisch	hydrophob	ungeladen	polar	2154
Valin	Val	V	aliphatisch	hydrophob	ungeladen	unpolar	4617
							$\Sigma = 58115$

2.1.2 Primärstruktur, Sekundärstruktur, Tertiärstruktur

Primärstruktur

Die lineare Abfolge der Aminosäuren in der Proteinkette und damit ihre kovalente Verknüpfung, wird Primärstruktur genannt. Sie sind über C_α -Atome mit jeweils einer Peptidbindung miteinander verknüpft, die den C-Terminus der vorherigen Aminosäure mit dem N-Terminus der nachfolgenden Aminosäure verbindet. Dabei werden ausschließlich Aminosäuren in L-Konformation verwendet. Das Rückgrat der Proteine besteht aus sich wiederholenden Grundeinheiten. Diese sind die C_α -Atome der Aminosäuren sowie die Peptidbindungen, welche benachbarte C_α -Atome miteinander verknüpfen. Dies ist in Abbildung 2 dargestellt.

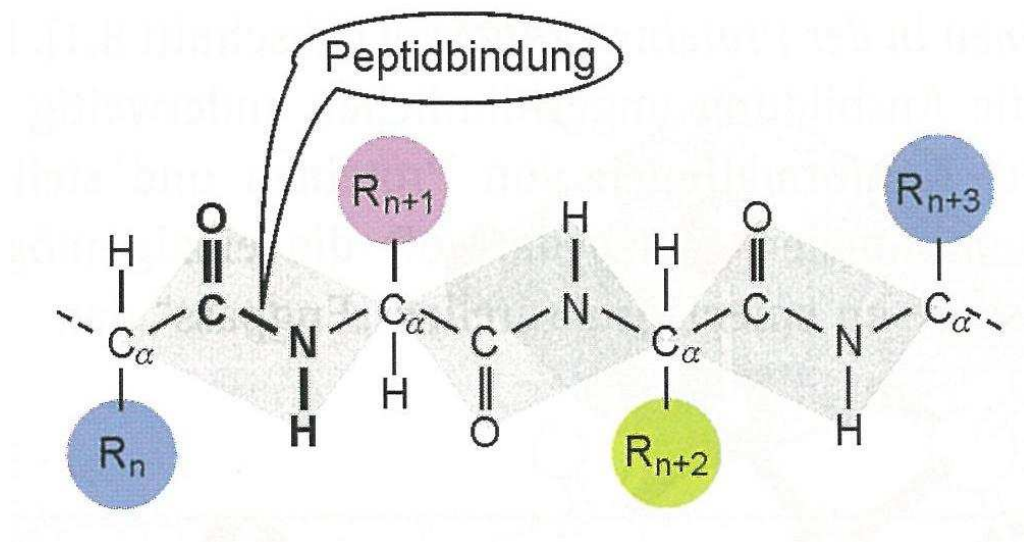


Abbildung 2: Schematische Darstellung eines Rückgrats eines Proteins. [6]

Da das freie Elektronenpaar am Stickstoff der Doppelbindung der Carbonylgruppe benachbart ist, kommt es zu einer Delokalisation der Elektronen über die Peptidbindung. Dieser Doppelbindungscharakter führt zur Aufhebung der freien Drehbarkeit um die Peptidbindung. Prinzipiell liegen die Peptidbindungen in der *cis*- oder *trans*-Konformation vor, wobei die *trans*-Form (die Substituenten *H* und *O* auf gegenüberliegenden Seiten der Peptidbindung) die energetisch begünstigte Konformation ist. Nur bei einer Peptidbindung einer beliebigen Aminosäure mit Prolin ist die *cis*-Form mit der *trans*-Form energetisch vergleichbar, was bei der Faltung bei Proteinen von Bedeutung ist. Ebenfalls von Bedeutung bei der Faltung ist Glycin, da es keine voluminöse Seitenkette, sondern nur ein kleines Wasserstoffatom besitzt. Deshalb nimmt auch Glycin wichtige Positionen in der Proteinstruktur ein. [6]

Sekundärstruktur

Die Ebene der Sekundärstruktur definiert die räumliche Organisation von benachbarten Aminosäuren in der linearen Sequenz ohne Berücksichtigung der Konformation der Seitenketten. Dabei entstehen drei maßgebliche Sekundärstrukturelemente von Proteinen: die α -Helix, das β -Faltblatt (engl. *β -sheet*) und die β -Schleife (engl. *Loop, Coil*).

Die α -Helix ist eine wendelförmige Struktur, die meist rechtsgängig ansteigt, wenn man entlang der Helixachse schaut. Wasserstoffbrücken, die zwischen Aminosäuren gebildet werden, die jeweils drei Reste voneinander getrennt liegen, also beispielsweise zwischen der dritten und siebten Aminosäure, stabilisieren diese Struktur. Diese Wechselwirkungen wiederholen sich periodisch und beziehen alle Aminosäuren der Helix ein. Die rechtsgängige α -Helix verfügt über charakteristische Kenngrößen. Jeder Rest ist um 100° relativ zu seinem Vorgänger um die Helixachse gedreht. Somit ergibt sich nach 3,6 Aminosäureresten eine komplette Helixwindung. Die Seitenketten der Aminosäuren stehen nahezu Senkrecht zur Helixachse und zeigen radial nach außen, wie Abbildung 3 verdeutlicht. Durch die helikale Struktur zeigen Donor und Akzeptor der Wasserstoffbrücken direkt aufeinander zu. Somit wird die Ausbildung von optimalen Wasserstoffbrücken begünstigt. Die meisten Aminosäuren können ihre Seitenkette problemlos in eine α -helikale Struktur unterbringen. Prolin stellt eine wichtige Ausnahme dieser Regel dar, weil dessen Aminogruppe in einem Ring aus fünf Atomen integriert ist. Ein in eine Polypeptidkette eingebautes Prolin besitzt kein Wasserstoffatom mehr und kann daher keine Wasserstoffbrücke zu einer CO-Gruppe ausbilden. Weiterhin verhindert der durch den Ringschluss eingeschränkte Spielraum um *Phi* und *Psi*, die Torsionswinkel um die Gerüstkonformation der Polypeptidkette, die Ausbildung einer idealen α -helikalen Struktur. Bei linksgängigen Helices würden sich Haupt- und Seitenkette zu nah kommen, sodass diese Struktur überwiegend aus Glycin und Prolin besteht, die, wie schon erwähnt, ungewöhnliche Konformationen ermöglichen. Linksgängige Helices finden sich im Kollagen. Der membranüberbrückende Teil von Membranproteinen besteht meist aus α -helicalen Strukturen [27].

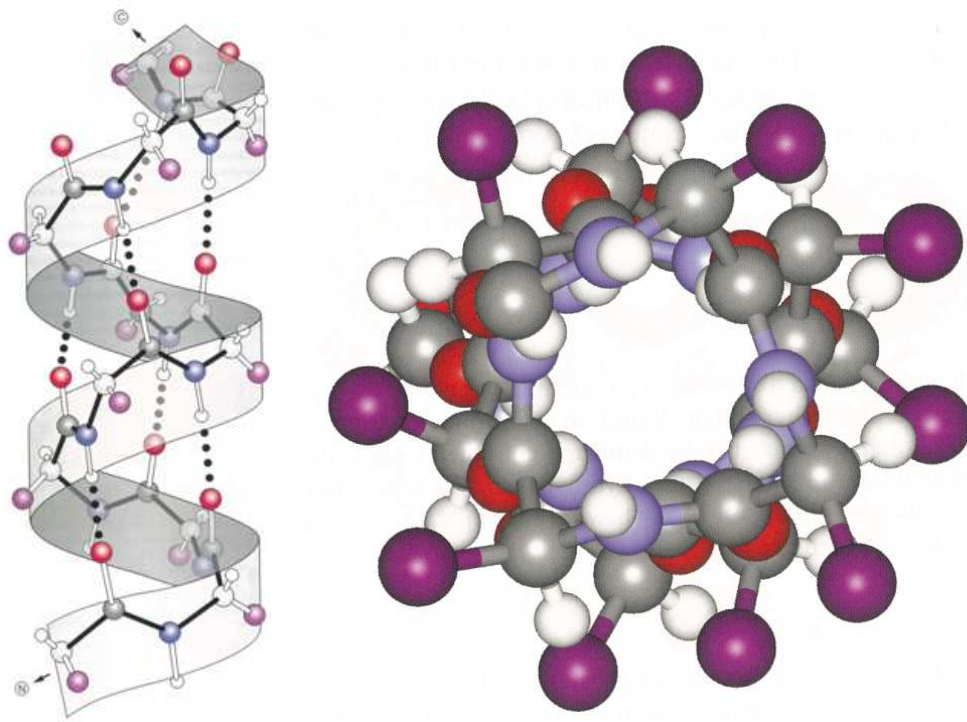


Abbildung 3: Darstellung der α -Helix.

Links ist die Seitenansicht und rechts die Draufsicht dargestellt. [6]

Ein weiteres wichtiges Sekundärstrukturelement ist das β -Faltblatt, welches aus mehreren β -Strängen (engl. β -strands) aufgebaut ist. Hierbei müssen, im Gegensatz zur α -Helix, die Stränge nicht aufeinanderfolgend an einer Polypeptidkette sein. Es können sogar unterschiedliche Abschnitte auf mehreren Polypeptidketten sein. Die beteiligten β -Stränge können dabei sowohl parallel, als auch antiparallel zueinander verlaufen. Weiterhin sind sie so nebeneinander angeordnet, dass sich Wasserstoffbrücken zwischen den CO- und NH-Gruppen benachbarter Stränge ausbilden können. β -Faltblatt Strukturen werden durch Isoleucin, Valin, Threonin, Tyrosin und Phenylalanin begünstigt. Prolinreste sind dagegen selten in Faltblättern zu finden.

Das dritte „klassische“ Sekundärstrukturelement ist die β -Schleife. Typischerweise verbindet sie zwei β -Strangsegmente, die ein antiparalleles Faltblatt bilden, über eine kurze Schleife. Zwei Varianten von Schleifen sind am gängigsten – Typ I und Typ II, die jeweils vier Reste umfassen. Beide Schleifentypen werden durch eine Wasserstoffbrücke zwischen dem ersten und vierten Rest stabilisiert. Die beiden eingeschlossenen Reste bilden über unterschiedliche Torsionswinkel zwei verschiedene Typen. Meist sind auf der Oberfläche von Proteinen β -Schleifen vertreten, wobei Prolin und Glycin am häufigsten darin vorkommen. [6] Abbildung 4 verdeutlicht eine mögliche Anordnung der drei Sekundärstrukturelemente in einem Protein.

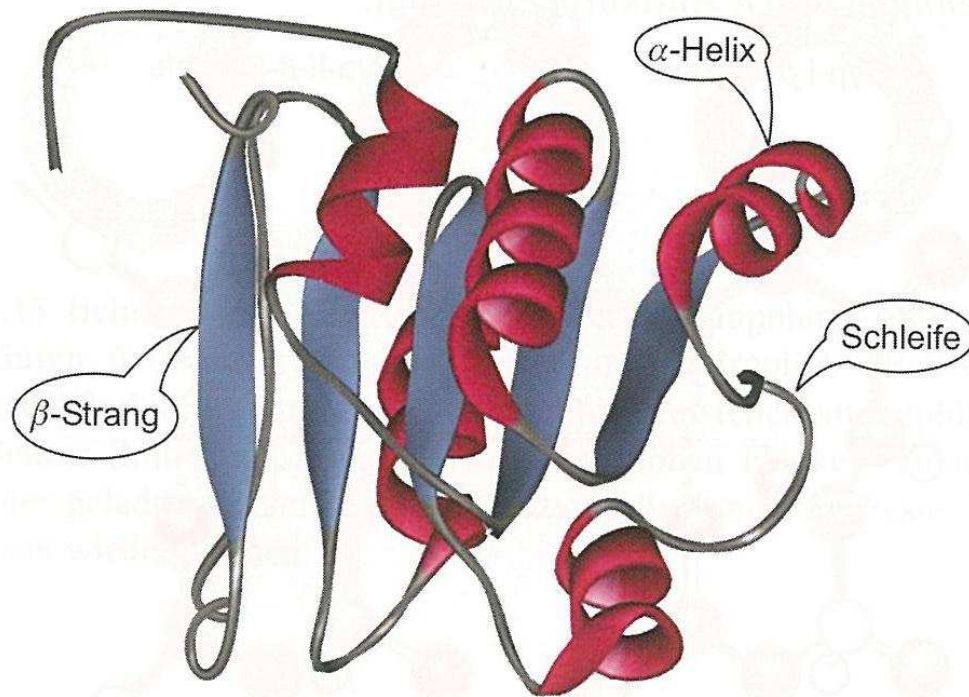


Abbildung 4: Ein schematisch dargestelltes Protein mit den Sekundärstrukturelementen. [6]

Tertiärstruktur

Die Tertiärstruktur besteht aus Sekundärstrukturelementen, die in erster Linie durch nichtkovalente Wechselwirkungen (ionische Bindungen, van-der-Waals-Kräfte, Wasserstoffbrücken) verknüpft sind. Da jede einzelne Bindung mindestens eine Größenordnung schwächer ist, als eine kovalente Bindung, ist das Zusammenwirken einer großen Zahl nichtkovalenter Kontakte für den Zusammenhalt der Proteinstruktur nötig. [6] Für deren Ausbildung ist der hydrophobe Effekt ein weiterer wichtiger Punkt. Dieser Effekt beruht auf der Tendenz hydrophober Moleküle oder Molekülbereiche, den Kontakt zu Wasser zu meiden. Die hydrophoben Moleküle neigen daher in einem wässrigen Umfeld zur Selbstaggregation, um die Kontaktfläche mit Wasser zu minimieren. Man nimmt an, dass das Wasser die Moleküle umschließt, da Wasser keine hydrophoben Moleküle in sein Netzwerk aus Wasserstoffbrücken einbinden kann. Dadurch erhöht sich die Ordnung des Wassers und seine Entropie wird verringert. Durch die Selbstaggregation hydrophober Gruppen wird der Entropieverlust des Wassers minimiert. Das erklärt, warum polare und geladene Seitenketten sich überwiegend auf der Oberfläche und umgekehrt sich hydrophobe Seitenketten dicht gepackt im Innenraum des Proteins befinden. Somit ist der Innenraum des Proteins wasserfrei. [11]

2.2 Interatomare Kräfte und Proteinstruktur

Nachdem nun bereits die Formen, wie sich ein Protein falten kann, angesprochen wurden, werden jetzt die interatomaren Kräfte und Wechselwirkungen, die an der Proteinfaltung beteiligt sind, behandelt.

2.2.1 Kovalente Wechselwirkungen

Sehr starke Kräfte, die Atome zu Molekülen verbinden, sind kovalente Wechselwirkungen. Sie besitzen die kürzeste Reichweite mit 0,1 nm bis 0,15 nm, sind aber in der Lage die chemische Natur zu verändern. Ursache dafür ist das Teilen der Elektronen in der Atomhülle. Es handelt sich danach nicht mehr um selbständige Einheiten, sondern um Teile in einem Molekülgefüge. Anschaulich ist dieser Sachverhalt in Abbildung 5 anhand von Chloratomen dargestellt. [12]

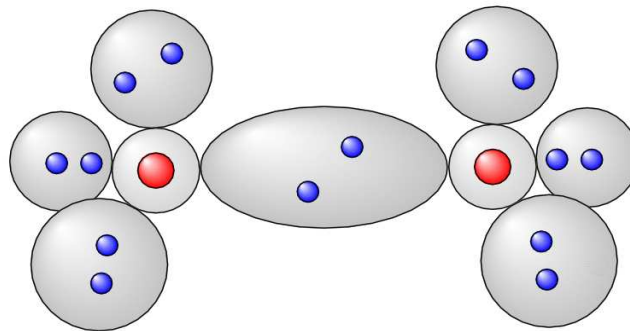


Abbildung 5: Molekülgefüge von Chloratomen. [13]

Somit werden Aminosäuren, die über Peptidbindungen miteinander verknüpft sind, streng genommen, Teil eines neuen, großen Gesamtmoleküls. Weiterhin herrschen durch die durch kovalente Bindungen miteinander verknüpften Atome, innerhalb der Proteinstruktur, starke Einschränkungen für die erlaubten Abstände zwischen den Atomen.

Für die Bildung von kovalenten Bindungen sind sehr eingeschränkte physikalisch-chemische Eigenschaften notwendig. Alle Bindungslängen und Winkel, besonders der Rotationswinkel um das C_α -Atom, dürfen nicht zu sehr vom erlaubten Wertebereich abweichen. Die Bindungslänge hängt von Art und Größe der beteiligten Atome und der Anzahl der von ihnen gemeinsam genutzten Elektronen ab. Die Bindung wird umso kürzer und stärker, je mehr Elektronen beteiligt sind. Die Bindungswinkel hingegen werden durch Elektronenwolken, die Atomorbitale, bestimmt. Sie sind schematisch in Abbildung 6 am Beispiel von Glycerin dargestellt. Sterische Einschränkungen bestimmen primär die Rotationswinkel um die zentrale Bindung zwischen zwei verbundenen Atompaaren. [12]

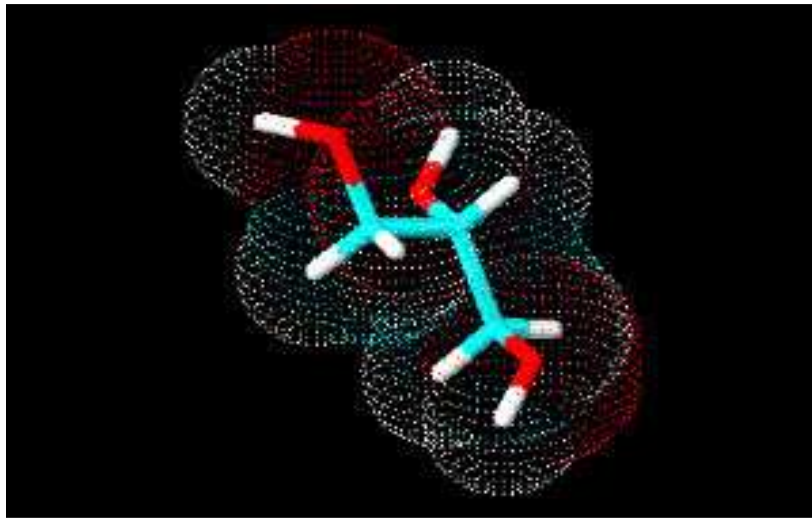


Abbildung 6: Molekülmodell Glycerin mit Elektronenwolken. [14]

2.2.2 Wasserstoffbrückenbindungen

Wasserstoffbrückenbindungen treten dort auf, wo Protonendonoren und Protonenakzeptoren interagieren. Ein Protonendonor ist in diesem Fall eine chemische Gruppe, in der ein Proton kovalent an ein stark elektronegatives Sauerstoffatom, Stickstoffatom oder Fluoratom gebunden ist. Somit ist die kovalente Bindung zwischen beiden Atomen polarisiert, positiv auf der Protonenseite und negativ auf der Stickstoff-, Sauerstoff-, oder Fluorseite. Ein Protonenakzeptor ist in diesem Fall ein elektronegatives Sauerstoff- Stickstoff- oder Fluoratom mit partiell negativem Ladungsüberschuss, jedoch ohne gebundenes Wasserstoffatom. Somit wird das positiv polarisierte Proton von der negativ polarisierten Akzeptorgruppe angezogen. Es entsteht eine nicht ganz so starke Bindung, wie die kovalente Bindung, aber sie ist kürzer und stärker als sonstige Bindungen. Die Stärke nimmt jedoch stark ab, wenn der Winkel zwischen den drei beteiligten Atomen vom Toleranzbereich Abweichungen zeigt. Innerhalb der Proteinstruktur gehören Wasserstoffbrückenbindungen zu den wichtigsten stabilisierenden Kräften. Am Proteinerückgrat befinden sich in regelmäßigen Abständen ein Protonenakzeptor, in Form des Sauerstoffatoms der Carboxylgruppe und ein Protonendonor, in Form der Aminogruppe. Die Haupttypen der Sekundärstruktur, die α -Helix und das β -Faltblatt, werden durch Wechselwirkungen dieser Gruppen stabilisiert. [12]

2.2.3 Hydrophobe und hydrophile Wechselwirkungen

Aminosäuren lassen sich in eine hydrophile und hydrophobe Teilklasse unterteilen. Alle in dieser Arbeit diskutierten Proteine sind Membranproteine, d. h. sie sind in Zellmembranen gebunden und stehen somit nur teilweise in ständigen Kontakt mit Wassermolekülen.

Dieses Wasser, als Lösungsmittel, besitzt Eigenschaften, die wesentlich dazu beitragen, dass Abschnitte von Membranproteinen, die aus der Membran in das Lösungsmittel – meist Wasser – hineinragen, in der für sie typischen kompakten Form vorliegen können. Die Polarität von Wassermolekülen führt dazu, dass jedes einzelne Wassermolekül im flüssigen Wasser vier Wasserstoffbrückenbindungen mit seinen Nachbarn ausbilden kann. Dieses ununterbrochene Gitter von Wassermolekülen, die durch Wasserstoffbrücken miteinander in Kontakt stehen, tragen zu den hohen Schmelz- und Siedetemperaturen des Wassers bei. Weiterhin resultieren geringe Kompressibilität und hohe Oberflächenspannung. Diese Eigenschaften nehmen wesentlichen Einfluss an der Interaktion von Wasser mit unlöslichen Proteinen. [12]

Ein in Wasser gelöstes, unpolares Molekül hingegen unterbricht das regelmäßige Wasserstoffbrückengitter des flüssigen Wassers. Denn anstatt nun frei in jede beliebige Lage rotieren zu können, müssen die Wassermoleküle um das unpolare Fremdmolekül herum einen Käfig bilden. Dieses Umorientieren ist in Bezug auf die Entropie ungünstig. Somit sind die Energetischen Kosten umso höher, je größer das unpolare, gelöste Molekül ist, weil sich mehr Wassermoleküle umorientieren müssen, um es abzuschirmen. Obwohl sie im Wesentlichen aus unpolaren Gruppen aufgebaut sind, existieren globuläre Proteine gelöst in Wasser. Damit sie überhaupt gelöst existieren können, sind sie gezwungen, auf der Moleküloberfläche Gruppen zu präsentieren, die mit dem Wasser Wasserstoffbrückenbindungen ausbilden können. Somit finden sich auf der Oberfläche von Proteinen bevorzugt Aminosäuren mit hydrophilen Seitenketten. [12]

2.2.4 Ladungs-Ladungs-, Ladungs-Dipol- u. Dipol-Dipol-Wechselwirkungen

Verschiedene wichtige Kräfte können zwischen polarisierten und/oder geladenen Atomen wirken. Geladene Atome sind Ionen, die aufgrund von Elektronenüberschuss oder Elektronendefizit positive oder negative Nettoladungen aufweisen. Die unterschiedlich geladenen Ionen werden mit einer Kraft angezogen, die sowohl vom inversen Abstand der Ionen, als auch von der Größe der Ladung abhängt.

In Proteinen kann es unter normalen physiologischen Bedingungen zwischen den Seitenketten von sauren und basischen Aminosäuren zu einem Austausch eines nicht fest gebundenen Protons kommen. Diese Salzbrücken, wie die in den Proteinen vorkommenden Ladungs-Ladungs-Wechselwirkungen genannt werden, tragen nicht unwesentlich zur Stabilisierung der Proteinstruktur bei.

Zwischen nicht-ionischen Gruppen und Ladungsträgern treten schwächere Wechselwirkungen auf. Moleküle mit Dipolcharakter gehören zu den nicht-ionischen Gruppen. Diese Dipolmoleküle besitzen immer ein Ende mit einem partiell positiven bzw.

ein Ende mit einem partiell negativen Ladungsüberschuss. Ein derartiger Dipol kann als Vektor angesehen werden, welcher die Stärke der Polarisierung der Atombindung angibt. Da viele Seitenketten von Aminosäuren und in ähnlichem Maße das Peptidrückgrat, starken Dipolcharakter besitzen, spielen diese Wechselwirkungstypen bei der Strukturstabilisierung eine entscheidende Rolle. [12]

2.2.5 Van-der-Waals-Kräfte

Van-der-Waals-Kräfte resultieren von den durch Zufallsfluktuationen in der atomaren Elektronenhülle spontan hervorgerufenen Dipolen. Daher handelt es sich um eine unspezifisch, aber relativ weitreichende, intermolekulare Kraft und somit spielen Van-der-Waals-Wechselwirkungen bei der Proteinfaltung und bei der Protein-Protein Interaktion eine signifikante Rolle. [12]

2.2.6 Abstoßende Kräfte

Sterische Wechselwirkungen gehören zu abstoßenden Kräften und besitzen nur eine kurze Reichweite. Die Stärke steigt allerdings stark an, wenn der Abstand zwischen zwei Atomen wesentlich kleiner wird. Dieser Sachverhalt ist in Abbildung 7 visualisiert. Der kritische Radius, bei dessen Unterschreitung die Abstoßungskräfte rapide ansteigen, wird durch eine Kugel um jedes Atomzentrum beschrieben. Es ist ein energetisch sehr ungünstiger Zustand, wenn zwei ungebundene Atome auf diese Art in Kontakt treten. Daher kommt ein solcher Zustand in realen Molekülen nicht vor. [12]

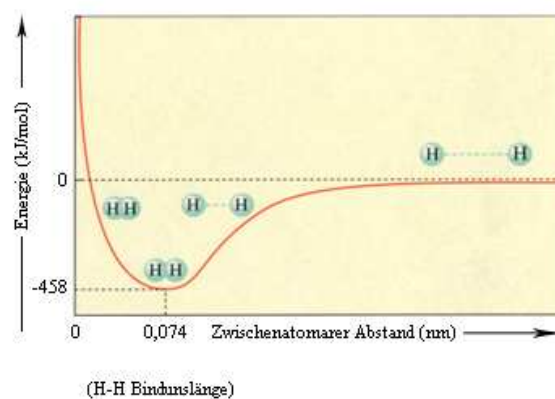


Abbildung 7: Darstellung sterischer Wechselwirkungen zwischen Atomen. [15]

2.3 Ein Einblick in die Proteinfaltung – Die Energielandschaft

Jedes Protein besitzt eine spezifische native Struktur, in der es wirksam werden kann. Um diese zu erreichen, muss es seinen nicht-nativen Zustand überwinden. Dabei tritt das Protein in vielen verschiedenen Konformationen auf und faltet sich innerhalb von Sekundenbruchteilen. Der heutige Wissensstand über die Proteinfaltung ist intuitiv geprägt, denn während der Faltung können auch bestimmte Zustände mehrmals auftreten. Die genaue Erklärung dieses Phänomens ist eine der größten Herausforderungen der modernen Wissenschaft und erfordert interdisziplinäre Kooperation mit Fachbereichen der Chemie, Informatik, Mathematik und Physik. Erste aufschlussreiche Ansätze zu diesem Problem formulierte *Cyrus Levinthal* bereits 1968. Wenn eine Aminosäurekette mit 100 Aminosäuren betrachtet wird und die Voraussetzung angenommen wird, dass jede Aminosäure in allen drei Sekundärstrukturen - Helix, Sheet und Coil - vorliegen kann, so gilt gemäß stochastischer Definition für eine Variation, in welcher Wiederholungen zugelassen werden, n^k . Auf unser Beispiel bezogen steht n für die Anzahl der Konformationen, somit drei, und k für die Anzahl der Aminosäuren, also 100. Das Ergebnis beschreibt nun näherungsweise 10^{48} verschiedene Konformationen. Unter Berücksichtigung der Tatsache, dass Rotationen um die Bindungen der Atome maximal 10^{14} -mal pro Sekunde geschehen können, wäre die Laufzeit etwa 10^{26} Jahre. [16] Man bedenke nun einmal das Alter des Universums und sehe selbst das Hindernis. Das Problem dieses Ansatzes ist, dass nicht alle Konformationen gleichwahrscheinlich auftreten können. Jene mit einer geringeren freien Energie sind wahrscheinlicher als Konformationen mit einer höheren. Dies erkannte auch *Levinthal* und er veranschaulichte seine Annahme, dass es spezifische Faltungsmuster in Proteinen gibt, mit der Darstellung eines Golflochs auf dem Rasen. Der Rasen bildet dabei das Protein in seinen ungefalteten Konformationen ab, wobei es einen bestimmten Faltungsweg (engl. *folding pathway*) gibt, welchen das Protein zur Ausbildung der nativen Struktur beschreibt. Diese Ausführung ist in Abbildung 8 dargestellt. [17]

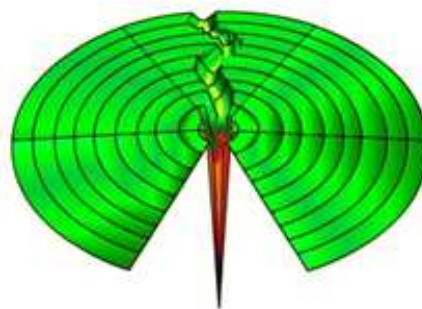


Abbildung 8: Erste Vorstellung des Faltungstrichters nach Levinthal. [16]

Die verschiedenen Zustände während der Faltung legen die Vermutung nahe, dass sich die unterschiedlichen Konformationen auch in der Energie des Proteins widerspiegeln. Der funktionale Zustand stellt die stabilste Form des Proteins dar, weshalb seine freie Energie im Vergleich zu allen instabilen Zwischenkonformationen am geringsten ist oder mindestens sehr nah am Minimum liegen muss. Der Wechsel der Konformationen erfolgt durch die dauerhaft vorherrschende *Brownsche Bewegung* des Moleküls. [18] Durch die sogenannte Energielandschaft (engl. *energy landscape*) nach [18] lassen sich die energetischen Änderungen optisch sehr ansprechend visualisieren. Sie zeigt eine verbesserte Darstellungsform von *Levinthal* auf. Die Proteinfaltung ist ein sehr komplexer Prozess, welcher bei konstantem Druck und Temperatur abläuft. Einbezogen werden auch strukturelle Veränderungen der Seitenketten der Aminosäuren sowie des Lösungsmittels. Das natürliche thermodynamische Potential zur Beschreibung eines Prozesses ist die *Gibbs-Energie*, benannt nach ihrem Begründer *Josiah Willard Gibbs*. Die detaillierteste Darstellung der Energielandschaft ist die Beschreibung als Funktion des Durchschnittswertes der freien Energien über die Lösungsmittel-Koordinaten. Diese Funktion wird auf jedes Atom des Proteins angewendet. Die freie Energie weist viele lokale Minima auf. Diese entsprechen oft Anregungsenergien in Verbindung mit individuellen lokalen Veränderungen in der Struktur, beispielsweise Rotationen von Seitenketten. [19] Abbildung 9 verdeutlicht die Vorstellung des Energietrichters mit der Energielandschaft am unteren Ende des Trichters. Verschiedene Konformationen einer zufälligen Aminosäuresequenz sind beispielhaft als Kugelgitter und die Energielandschaft ist vergrößert dargestellt.

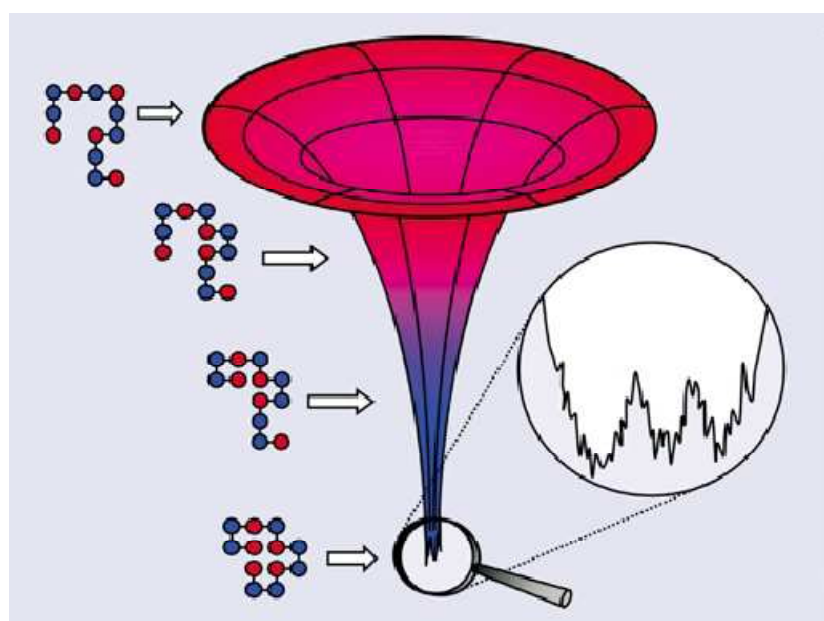


Abbildung 9: Schematische Darstellung des Faltungstrichters eines Proteins. [18]

Der Faltungstrichter ist im oberen Bereich sehr breit ausgeprägt, der eine sehr große Anzahl verschiedener Wege der Faltung verdeutlicht. Wenn ein bestimmter Faltungsweg durch eine Energiebarriere gehemmt ist, wird mit hoher Wahrscheinlichkeit eine Alternativroute eingeschlagen. Im unteren Bereich ist die starke Verengung des Trichters ersichtlich, die das Protein in seinen Konformationsänderungen hin zum nativen Zustand stark limitiert. Das Protein muss den Pfad zur nativen Form relativ zielsicher finden. Es kann dabei auch zu Wiederholungen von bereits durchlaufenen Konformationen kommen, d. h. das Protein entfernt sich wieder von seinem energetischen Minimum. Proteine in nativer Konformation haben zusätzlich den Drang, sich energetisch optimal zu ihren Nachbarn anzuordnen. Die Darstellung der Oberfläche der freien Energien ist ein entscheidendes Charakteristikum. Diese ist anschaulich bereits in Abbildung 9 dargestellt. In nachstehender Abbildung 10 sind beide auftretenden Grundformen der Energielandschaft noch einmal genauer visualisiert. [18]

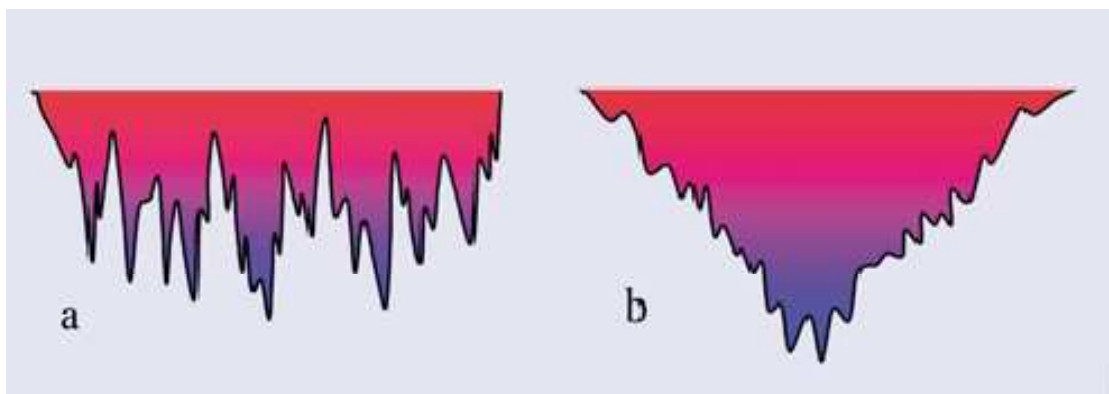


Abbildung 10: Grundformen der Energielandschaften.

Eine frustrierte Aminosäuresequenz (a) und ein natürlich vorkommendes Protein (b). [18]

Die schematische Energielandschaft *a* definiert eine frustrierte Aminosäuresequenz, *b* die eines Proteins im nativen Zustand. Die Bezeichnung „frustriert“ sagt hierbei aus, dass es sich um eine zufällige Aminosäureabfolge handelt. Diese weisen eine raue, unstrukturierte und von vielen Minima geprägte Energielandschaft auf. In dem Minima-Zustand kann das Protein noch viele weitere Konformationssubzustände ausbilden, welche in ihrer Funktion oft sehr verschieden sind, aber nur leichte Strukturunterschiede aufweisen. Diese werden durch die vielen lokalen Minima entlang der Struktur dargestellt. Durch die Natur geschaffene Proteine beschreiben hingegen eine flache, in Trichterform zulaufende Landschaft. Kleine Peaks deuten auf Energiebarrieren hin, welche jedoch sehr gering sind und die Faltung nur in einem geringen Maß beeinflussen. [18]

Abbildung 11 zeigt die Auftragung der Energielandschaft in ein Energie-Sequenz-Diagramm. Auch in diesen beiden Diagrammen erkennt man deutlich den Unterschied zwischen einer natürlichen und synthetischen Aminosäuresequenz. Im oberen Bereich ist eine flache Energielandschaft ersichtlich und im unteren eine raue Energielandschaft.

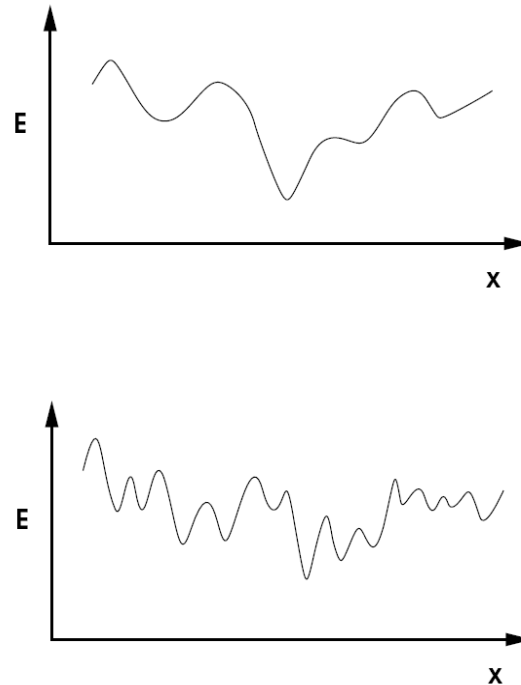


Abbildung 11: *Energieprofile in einem Energie-Sequenz-Diagramm. [19]
Die Y-Achse definiert die Energie des Systems, die X-Achse repräsentiert jedes einzelne Atom der Aminosäuren des Proteins.*

Die Proteinfaltung ist bis zum heutigen Zeitpunkt trotz aufwendiger Computersimulationen noch nicht komplett vorhersagbar bzw. nur auf sehr kurze Aminosäuresequenzen übertragbar.

3 Energiefunktion bei Membranproteinen

In Proteinen ist der Grundzustand meist nur durch eine relativ kleine Energiebarriere von dem Ensemble entfalteter Strukturen getrennt ($\sim 20 \text{ k}_\text{B}T$, entspricht ~ 10 Wasserstoffbrückenbindungen) [20]. Strukturell unterscheiden sich die entfalteten Zustände allerdings erheblich vom nativen Zustand. Somit musste die maximale Reichweite und Form der Wechselwirkung in einem diskretisierten Modell sorgfältig gewählt werden. Interaktionen zwischen Residuen müssten einerseits so unabhängig von kleinen Strukturabweichungen des Modells gegenüber dem nativen Zustand sein, dass diese nicht zu unterschiedlichen Energien führen. Andererseits musste der Unterschied zwischen Grundzustand und entfalteten Zuständen korrekt erfasst werden. Dies wurde durch ein Wallpotential erreicht. Die Wechselwirkungsreichweite im Modell wurde 8 \AA gewählt, sodass sie, angelehnt an [21, 22], etwas größer als die maximale Wechselwirkungsreichweite in Proteinen ist. [22]

Unterschiede in den Wechselwirkungen zwischen Aminosäuren sind auf deren Seitenketten zurückzuführen, da sich die einzelnen Aminosäuren nur in diesen unterscheiden. Die Wechselwirkungen eines Proteins lassen sich in Interaktionen der Aminosäuren mit der Umgebung und in Interaktionen mit anderen Residuen aufteilen. [22]

3.1 Wechselwirkungen mit der Umgebung

Die Möglichkeit von Aminosäuren mit dem umgebenden Lösungsmittel zu wechselwirken beruht auf der Lösungsmittelzugänglichkeit der Oberfläche von Proteinen. Diese Lösungsmittelzugänglichkeit von Proteinoberflächen kann effizient vom POPS-Algorithmus berechnet werden [23]. Er errechnet die *solvent accessible surface area* (SASA) jeder Aminosäure in \AA^2 . Um diesen Ansatz umzusetzen wurde das Innen/Außen-Kriterium¹ aus [22] erweitert und an Membranproteine angepasst. Ist die Lösungsmittelzugänglichkeit gering, gilt eine Aminosäure als *innen*. Diese Aminosäuren orientieren sich vermehrt zum Zentrum des Proteins. Ist die Lösungsmittelzugänglichkeit hoch, gilt eine Aminosäure als *außen*. Solch bestimmte Aminosäuren orientieren sich zur Oberfläche des Proteins. Die Ausgabe des POPS-Algorithmus wurde für jedes Protein in einer eigenen Datei, dargestellt in nachstehender Abbildung 12, gespeichert.

¹ Das in [22] erstellte Innen/Außen-Kriterium unterschied nicht, ob Aminosäuren außerhalb der Membran oder innerhalb der Membran vorkommen. Für Aminosäuren in globulären Proteinen existieren zwei Zustände, für Aminosäuren in Membranproteinen existieren vier Zustände des Innen/Außen-Kriteriums.

```

=== ATOM SASAS ===
AtomNr  AtomNe  ResiNe  Chain  ResidNr  SASA/A^2  N(overlaps)  Atom Group
1        N      SER     P       71       5.49      28           3
2        CA     SER     P       71      10.62     24           5
3        C      SER     P       71       2.84     24           3
4        O      SER     P       71      22.32     25           3
5        CB     SER     P       71      40.05     26           5
6        OG     SER     P       71      15.35     26           3
7        N      GLY     P       72       9.93     26           3
8        CA     GLY     P       72      26.38     36           5
9        C      GLY     P       72       0.65     45           3
10       O      GLY     P       72       4.42     47           3
11       N      PHE     P       73       3.62     38           3
12       CA     PHE     P       73       2.35     44           5
13       C      PHE     P       73       0.92     39           3
14       O      PHE     P       73      10.97     28           3
15       CB     PHE     P       73       7.02     47           5
16       CG     PHE     P       73       1.19     53           4
17       CD1    PHE     P       73       0.68     66           4
18       CD2    PHE     P       73       1.67     55           4
19       CE1    PHE     P       73       0.92     65           4
20       CE2    PHE     P       73       1.39     57           4
21       CZ     PHE     P       73       1.24     62           4
22       N      GLU     P       74       1.13     40           3
23       CA     GLU     P       74       1.47     47           5
24       C      GLU     P       74       0.10     55           3
25       O      GLU     P       74       1.83     56           3
26       CB     GLU     P       74       3.50     54           5
27       CG     GLU     P       74       9.18     48           5
28       CD     GLU     P       74       4.68     32           3
29       OE1    GLU     P       74      12.69     27           2
30       OE2    GLU     P       74      20.18     19           2
31       N      PHE     P       75       0.78     56           3
32       CA     PHE     P       75       0.58     61           5
33       C      PHE     P       75       0.21     57           3
34       O      PHE     P       75       2.44     47           3
35       CB     PHE     P       75       2.38     56           5

```

Abbildung 12: Die „POPS area per atom“ Ausgabe eines Proteins in einer Datei gespeichert. Der POPS-Algorithmus unterteilt die Residuen in einzelne Atome bzw. Atomgruppen und berechnet jeweils die $SASA/\text{\AA}^2$. In jeder Zeile befindet sich jedes Atom bzw. jede Atomgruppe einer Aminosäure. Zusätzlich sind unter anderem eine laufende Atomnummer, der 3-letter-Code, die Ketten-ID, der Index der Aminosäure in dem PDB-File und der $SASA/\text{\AA}^2$ -Wert angegeben.

Für die Bestimmung der Orientierung einer Aminosäure mussten die Angaben, die der POPS-Algorithmus berechnet hat, ausgewertet werden. Dazu wurde entsprechend der Größe jeder Aminosäure ein *Längenwert* eingeführt, der in Tabelle 2 visualisiert wurde. Der *Längenwert* verhält sich äquivalent zur Anzahl der vom POPS-Algorithmus ausgegebener Atome bzw. Atomgruppen pro Aminosäure.

Tabelle 2: Längenwerte der einzelnen Aminosäuren.

Längenwert	4	5	6	7	8	9	10	11	12	14
Aminosäuren	Gly	Ala	Cys	Pro	Asn	Gln	His	Arg	Tyr	Trp
			Ser	Thr	Asp	Glu		Phe		
				Val	Ile	Lys				
					Leu					
					Met					

Zusätzlich wurden ein *Score* sowie drei Schwellwerte (3,75; 10; 13) eingeführt. Nacheinander wurden alle *SASA-Werte* jedes Atoms/ jeder Atomgruppe betrachtet. Der *Score* wurde für jede Aminosäure neu ermittelt. Er wird:

- um eins erhöht, falls der *SASA-Wert* mindestens 3,75 beträgt;
- um zwei erhöht, falls der *SASA-Wert* mindestens 10,00 beträgt;
- um fünf erhöht, falls der *SASA-Wert* mindestens 13,00 beträgt.

Eine Aminosäure gilt als *außen*, wenn eine der folgenden fünf Bedingungen erfüllt ist:

- Der Quotient aus *Score* und *Längenwert* einer Aminosäure beträgt mindestens 0,5.
- Der *Gesamtlöslichkeitswert*² beträgt mindestens 18,5 und der Quotient aus *Score* und *Längenwert* dieser Aminosäure beträgt mindestens 0,45.
- Der *Gesamtlöslichkeitswert*² beträgt mindestens 23,75 und der Quotient aus *Score* und *Längenwert* dieser Aminosäure beträgt mindestens 0,3.
- Der *Gesamtlöslichkeitswert*² beträgt mindestens 28 und der *Score* mindestens 3.
- Der *Gesamtlöslichkeitswert*² beträgt mindestens 40.

Ist keine der obigen Bedingungen erfüllt, gilt die Aminosäure als *innen*. Zusätzlich wird unterschieden, ob sich eine Aminosäure außerhalb der Membran, oder innerhalb der Membran befindet. Die einzelnen Häufigkeiten aller Aminosäuren im Datensatz unter diesem Kriterium sind in Tabelle 3 zusammengefasst.

Unter der Annahme, dass der Zustand einer Aminosäure durch diese Häufigkeiten charakterisiert ist, lässt sich mit Hilfe der Boltzmannverteilung die Energiedifferenz zwischen den beiden Zuständen berechnen:

$$e'_{i0} = -k_B T \ln \left(\frac{n_{in}}{n_{out}} \right) \quad (1)$$

n_{in} und n_{out} bezeichnet die Anzahl der Fälle, in denen die Aminosäure i im verwendeten Datensatz *innen*, bzw. *außen* ist - unter Berücksichtigung, ob sich Aminosäure i innerhalb oder außerhalb der Membran befindet. [22]

² Alle *SASA-Werte* aller Atome/Atomgruppen einer Aminosäure zusammengefasst.

Tabelle 3: Häufigkeiten der Zustände der einzelnen Aminosäuren für Membranproteine. Aminosäuren von Membranproteinen können vier Zustände im Innen/Außen-Kriterium annehmen. Sie können innerhalb der Membran als „außen“ oder „innen“ eingeordnet werden und sie können außerhalb der Membran als „außen“ oder „innen“ eingeordnet werden.

Aminosäure	innerhalb der Membran		außerhalb der Membran	
	innen	außen	innen	außen
Alanin	893	1582	1110	1076
Arginin	155	1502	124	328
Asparagin	226	1254	195	344
Asparaginsäure	291	1398	111	205
Cystein	181	161	140	149
Glutamin	129	1018	124	254
Glutaminsäure	205	1559	119	196
Glycin	710	1674	1089	722
Histidin	91	533	106	153
Isoleucin	616	1019	605	1403
Leucin	803	1804	790	2270
Lysin	115	1703	56	271
Methionin	178	494	230	397
Phenylalanin	544	818	588	1004
Prolin	264	1322	245	300
Serin	554	1468	667	529
Threonin	480	1353	461	628
Tryptophan	158	330	143	361
Tyrosin	347	759	307	519
Valin	763	1202	715	1402
Gesamt	7703	22953	7925	12511

3.2 Wechselwirkungen innerhalb des Proteins

Die Wechselwirkungen innerhalb des Proteins können analog zu Formel (1) folgendermaßen beschreiben werden:

$$e_{ij}^* = -k_B T \ln \left(\frac{n_{ij}}{N_{\text{contact}} p_i p_j} \right) \quad [22] \quad (2)$$

Dabei bezeichnet n_{ij} die Anzahl der beobachteten Kontakte zwischen den Aminosäuren i und j , N_{contact} die Gesamtzahl der Kontakte zwischen den Aminosäuren aller verwendeten Proteine und p_i die Wahrscheinlichkeit, dass ein zufällig gewähltes Residuum eines Proteins der Aminosäure i entspricht (in Tabelle 4 dargestellt). Der Nenner gibt die erwartete Anzahl spezifischer Aminosäure-Aminosäure-Kontakte an, falls die Wechselwirkungsenergie zwischen den jeweiligen Aminosäuren 0 wäre. [22]

Tabelle 4: Häufigkeiten der einzelnen Aminosäuren im Datensatz.

Wahrscheinlichkeit p_i , dass ein zufällig gewähltes Residuum der Aminosäure i entspricht. [22]

Aminosäure	p_i	Aminosäure	p_i	Aminosäure	p_i	Aminosäure	p_i
Ala	9,10 %	Arg	4,14 %	Asn	3,87 %	Asp	3,87 %
Cys	1,21 %	Gln	2,99 %	Glu	4,04 %	Gly	8,21 %
Ile	7,20 %	Leu	11,14 %	Lys	4,10 %	His	1,74 %
Met	2,60 %	Phe	5,92 %	Pro	4,22 %	Ser	6,27 %
Thr	5,75 %	Trp	1,99 %	Tyr	3,71 %	Val	7,94 %

3.3 Ein Paarpotential zur Beschreibung wechselwirkender Aminosäuren

Die endliche Ausdehnung von Aminosäuren kann maßgeblichen Einfluss auf die Strukturentstehung haben. [24] Um diese zu berücksichtigen, wurde für den Bereich zwischen 0 und 3 Å ein unendlich stark abstoßendes Potential zwischen zwei Residuen i und j verwendet. Dabei wurden nur die Abstände r_α und r_β ihrer C_α - und C_β -Atome betrachtet: [22]

$$E_{sterisch}(i,j) = \begin{cases} \infty: & r_{\alpha,ij} \leq 3 \vee r_{\alpha/\beta,ij} \leq 3 \vee r_{\beta,ij} \leq 3 \\ 0: & \text{else} \end{cases} \quad (3)$$

In Formel (1) werden keine absoluten Energien, sondern nur die Energiedifferenzen zwischen zwei Zuständen einer Aminosäure i behandelt. Diese Differenz wird durch den Unterschied zwischen *Innen* und *Außen* hervorgerufen, der sich in einer Differenz α_i in der Anzahl der zu Aminosäure i räumlich benachbarten Aminosäuren j mit $(r_{ij} \leq 8 \text{ Å})$ ausdrückt. Damit kann die Energie pro Kontakt mit Aminosäure i berechnet werden [22]:

$$e_{i0} = \frac{1}{\alpha_i} e'_{i0}. \quad (4)$$

Formel (1) - (4) kann anschließend zu einem Paarpotential zusammengefasst werden:

$$e_{ij} = e_{i0} + e_{j0} + e_{ij}^*. \quad (5)$$

(5) wird für die Berechnung aller Membranproteine in dieser Arbeit verwendet. [22]

3.4 Ergänzende Information zur Energiefunktion

Die üblichste physikalische Maßeinheit der Energie ist das Joule. Auf eine Umrechnung³ in J oder kcal/mol⁻¹ wird in dieser Arbeit jedoch verzichtet. Dies wird legitimiert, indem neben der Boltzmann-Konstante k_B auch die Temperatur T als konstant angenommen wird. $K_B T$ ist folglich ein Proportionalitätsfaktor, der den direkten Zusammenhang zwischen Formel (1) bzw. Formel (2) und der Energie in Joule beschreibt. Er kann vernachlässigt werden, ohne dass die physikalische Bedeutung von e_{ij} verloren geht. Die in dieser Arbeit erwähnten Energiewerte sind demnach einheitslos. Bei allen in dieser Arbeit berechneten Energiewerten handelt es sich um die freie Energie einzelner Aminosäuren, oder eines gesamten Proteins.

Das energetische Spektrum von Membranproteinen reicht von -10 bis 70.

3 Die Umrechnung ist durch die Multiplikation mit $K_B T$ zu realisieren.

4 Evaluierung der Energiefunktion

4.1 Erstellung meines Datensatzes

Die *Protein Data Bank of Transmembrane Proteins* (PDBTM) ist eine Datenbank, die vom *Institute of Enzymology* in Budapest produziert wird [25]. In dieser Datenbank sind Transmembrane Proteine hinterlegt, u. a. mit den Angaben, welcher *Region* eine betrachtete Aminosäure eines Proteins angehört. Diese *Regionen* sind in Abbildung 13 in einem Auszug dargestellt. Anhand dieser *Region* lässt sich ableiten, welche Aminosäuren außerhalb bzw. innerhalb der Membran auftreten. Für den Algorithmus zur Berechnung von Energieprofilen von Membranproteinen ist diese Information grundlegend wichtig.

```

<CHAIN CHAINID="P" NUM_TM="18" TYPE="beta">
  <SEQ>
    SGFEFHGYAR SGVIMNDSSA STKSGAYITP AGETGGAIGR LGNQADTYVE MNLEHKQTLTD NGATTRFKVM
    VADGQTSYND WTASTSDLNV RQAFVELGNL PTFAGPFKGS TLWAGKRFRD DNFDIHWIDS DVVFLAGTGG
    GIYDVKWNDDG LRSNFSLYGR NFGDIDDSSN SVQNYILTMN HFAGPLQMMV SGLRAKDNDK RKDSNGNLAK
    GDAANTGVHA LLGLHNSFY GLRDGSSKTA LLYGHGLGAE VKGIGSDGAL RPDADTWRIA SYGTTPLSEN
    WSVAPAMLAQ RSKDRYADGD SYQWATFNLR LIQAINQNFA LAYEGSYQYM DLKPEGYNDR QAVNGSFYKL
    TFAPTFKVGSG IGDFFSRPEI RFYTSWMDWS KKLNNYASDD ALGSDGFNSG GEWSFGVQME TW
  </SEQ>
  <REGION seq_beg="1" pdb_beg="71" seq_end="6" pdb_end="76" type="2"/>
  <REGION seq_beg="7" pdb_beg="77" seq_end="15" pdb_end="85" type="B"/>
  <REGION seq_beg="16" pdb_beg="86" seq_end="34" pdb_end="104" type="1"/>
  <REGION seq_beg="35" pdb_beg="105" seq_end="46" pdb_end="116" type="T"/>
  <REGION seq_beg="47" pdb_beg="117" seq_end="53" pdb_end="123" type="B"/>
  <REGION seq_beg="54" pdb_beg="124" seq_end="67" pdb_end="137" type="2"/>
  <REGION seq_beg="68" pdb_beg="138" seq_end="83" pdb_end="153" type="B"/>
  <REGION seq_beg="84" pdb_beg="154" seq_end="85" pdb_end="155" type="1"/>
  <REGION seq_beg="86" pdb_beg="156" seq_end="95" pdb_end="165" type="B"/>

```

Abbildung 13: XML-Auszug eines Proteins aus der PDBTM

Jede Proteinkette enthält eine oder mehrere „REGION“ Elemente, welche den Abschnitt der Proteinkette im Raum relativ zu der Membran lokalisieren. Der Typ der Region kann 1, 2, B, H, C, I, L und U sein. Dies steht für „Seite eins“, „Seite zwei“, „beta-Strand“, „alpha-Helix“, „Coil“, „Membran-innen“, „Membran-Loop“ und „unbekannt“. „Seite eins“ und „Seite zwei“ beziehen sich auf die beiden außen liegenden Seiten der Membran. Die Berechnung erfolgt nur anhand von PDB-Files. Daher ist eine genaue Zuordnung, welche Seite extrazellulär, bzw. intrazellulär ist, nicht möglich. Membran-inside ist der innere Teil eines beta barrel. Membrane-Loop entspricht dem Teil eines Proteins, welches die Membran nicht durchquert – nur eintaucht. [25]

Der Datensatz beruht auf einer Liste von nicht redundanten Proteinketten, die auf der PDBTM hinterlegt sind. Zum validieren dieser Proteine wurde ein Abgleich mit dem jeweiligen PDB-File, und dem zugehörigen Eintrag in der PDBTM vorgenommen. Falls ein Abschnitt innerhalb der Sequenz einer Proteinkette keiner *Region* zugeordnet werden konnte (*Region: unbekannt*), wurde diese Proteinkette aus dem Datensatz entfernt.

Ist eine unbekannte *Region* am Anfang oder Ende einer Sequenz vorgekommen, wurde diese Kette im Datensatz behalten. Grund dafür ist, dass zum einen diese unbekannten Abschnitte an den Enden des Proteins nicht in die Statistik eingreifen und zum anderen sich heraus stellte, dass es sich meist um Abschnitte handelt, die nicht zu der in den PDB-Files angegebenen Sequenz gehören. Dies ist an einem Beispiel eines Proteins im Datensatz in nachstehender Abbildung 14 dargestellt.

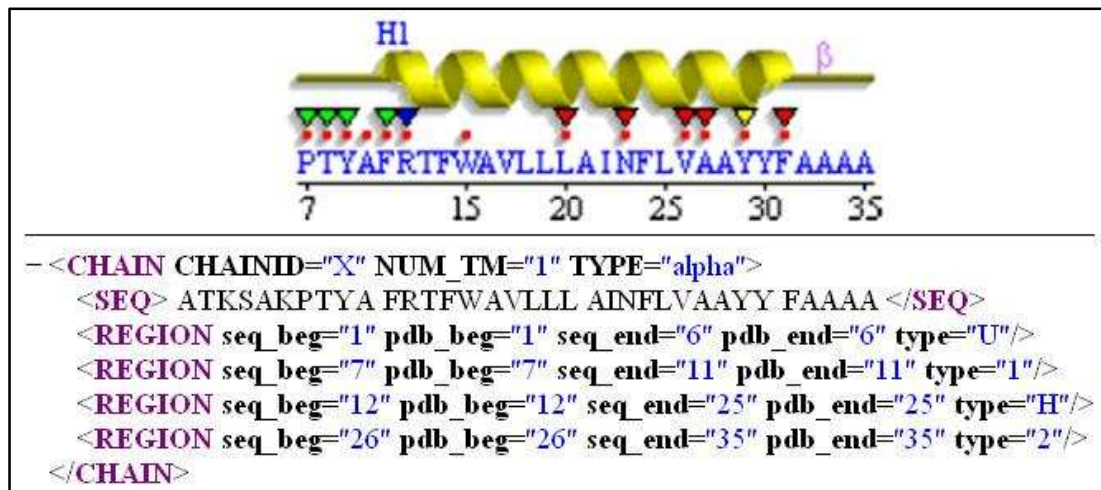


Abbildung 14: Vergleich von 1bj0 in der PDBTM und der PDBsum.

Der obere Teil zeigt den Abschnitt der Kette X, die in der PDB-File vollständig aufgeklärt ist. Der untere Teil zeigt einen XML-Abschnitt der Kette X des Proteins 1bj0, wie er in der PDBTM hinterlegt ist. Die gesamte bekannte Sequenz ist enthalten, jedoch auch einige Residuen, deren Lage im Raum unbekannt ist. [25], [26]

Um die Wechselwirkung von Aminosäuren mit dem umgebenden Lösungsmittel berechnen zu können, ist von jedem Protein im Datensatz der Output vom POPS-Server notwendig. Bei der Berechnung wurde ein *probe radius* von 1,4 Å und das Ausgabeformat *POPS area per atom* eingestellt [25]. Nicht zu jedem Protein im Datensatz konnte der POPS-Algorithmus eine erfolgreiche Berechnung durchführen. Aus diesem Grund ist der Datensatz, mit dem die Statistik für die Wechselwirkung mit der Umgebung erstellt wurde, um wenige Proteine kleiner, als der gesamte Datensatz, der in Tabelle 1 vorgestellt wurde.

4.2 Validierung des Berechnungsalgorithmus

Der ursprüngliche Algorithmus, der von Dr. Dressel im Rahmen seiner Dissertationsschrift entwickelt wurde, ist nicht geeignet Energieprofile von Membranproteinen zu berechnen. Der Unterschied zwischen beiden Algorithmen liegt in der Berechnung der Wechselwirkung der Aminosäuren mit der Umgebung. Das ursprünglich entwickelte Innen/Außen-Kriterium unterschied nicht zwischen Aminosäuren, die innerhalb der Membran bzw. außerhalb der Membran vorliegen.

Das in dieser Arbeit neu entwickelte und das ursprüngliche Innen/Außen-Kriterium wurden an einem Set⁴ aus globulären Proteinen getestet und anschließend miteinander verglichen. Das Ergebnis ist in nachstehender Tabelle 5 dargestellt. Von allen 20 Aminosäuren haben 17 die gleiche Tendenz, d. h. sie wurden von beiden Kriterien überwiegend innen oder überwiegend außen eingeordnet. Histidin, Methionin und Tyrosin wurden mit einer unterschiedlichen Tendenz von beiden Kriterien eingeteilt. 13 Aminosäuren haben eine Differenz zwischen beiden Quotienten beider Kriterien, die kleiner oder gleich 0,33 ist. Im Vergleich zu allen Aminosäuren haben Phenylalanin, Tryptophan, Tyrosin und Isoleucin eine relativ große Abweichung zwischen beiden Quotienten mit $\sim 1,0$. Die größte Abweichung mit 1,5 ist bei Methionin zu finden. Zwischen beiden Kriterien ergibt sich ein Korrelationskoeffizient von 0,8997, der zusätzlich unterstreicht, dass sie als annähernd gleich angesehen werden können.

Tabelle 5: Vergleich des ursprünglichen mit dem neu entwickelten Innen/Außen-Kriterium. Von den 20 kanonischen Aminosäuren sind die Quotienten aus der Gesamtanzahl der vorkommenden Aminosäuren „innen“ und der Gesamtanzahl der vorkommenden Aminosäuren „außen“ dargestellt.

Aminosäure	ursprünglich	neu entw.	Aminosäure	ursprünglich	neu entw.
A Ala	1,13	1,17	L Leu	2,83	2,51
R Arg	0,44	0,11	K Lys	0,36	0,07
N Asn	0,47	0,23	M Met	2,54	1,04
D Asp	0,42	0,29	F Phe	3,67	2,73
C Cys	2,45	2,92	P Pro	0,48	0,28
Q Gln	0,53	0,21	S Ser	0,53	0,57
E Glu	0,38	0,18	T Thr	0,65	0,49
G Gly	0,52	0,74	W Trp	2,90	1,94
H His	1,02	0,36	Y Tyr	2,06	0,99
I Ile	3,58	2,51	V Val	2,10	1,90

⁴ Das Set bestand aus ca.100 globulären Proteinen, die keine Liganden, nur eine Kette und eine Resolution von 2,0 Å oder besser bei der Strukturaufklärung durch X-Ray oder NMR vorzuweisen hatten.

Zur genaueren Betrachtung wurde auch ein Vergleich beider Innen/Außen-Kriterien an einem Protein vorgenommen. Die Wahl fiel dabei auf das Protein mit der PDB-ID 1a3h, da es sich um ein größeres Protein aus dem Set an globulären Proteinen handelte. Die Ergebnisse aus dem einzelnen Test sind in Tabelle 6 visualisiert. 80 % aller Residuen des Proteins wurden von beiden Kriterien gleich eingeordnet. Von allen 20 Aminosäuren haben bei diesem Protein elf die gleiche Tendenz. Zu diesen gehören unter anderem Alanin, Leucin und Lysin. Arginin und Isoleucin gehören zu den sechs Aminosäuren, die bei diesem Protein exakt gleich verteilt sind. Die einzigen Aminosäuren, die eine unterschiedliche Tendenz zwischen beiden Kriterien aufweisen sind Glycin und Histidin. Histidin unterscheidet sich zwischen beiden Kriterien aber nur um eine Aminosäure, somit ist Glycin die einzige Aminosäure, die von beiden Kriterien bei dem Protein mit der PDB-ID 1a3h unterschiedlich eingeordnet wurde. Cystein kommt im gesamten Protein nicht vor.

Tabelle 6: Vergleich beider Innen/Außen-Kriterien an einem Protein.

Die Werte geben die Anzahl der Aminosäuren in der jeweiligen Orientierung zur Oberfläche des Proteins (out) oder zum Inneren des Proteins (in) an. Gegenübergestellt sind hierbei das ursprüngliche sowie das neu entwickelte Kriterium.

Amino- säure	neu entw.		ursprünglich		Amino- säure	neu entw.		ursprünglich	
	in	out	in	out		in	out	in	out
A Ala	21	3	19	5	L Leu	11	5	14	2
R Arg	4	4	4	4	K Lys	4	6	3	7
N Asn	9	11	6	14	M Met	6	1	6	1
D Asp	8	17	4	21	F Phe	9	1	7	3
C Cys	0	0	0	0	P Pro	3	8	2	9
Q Gln	1	11	2	10	S Ser	14	6	13	7
E Glu	5	20	4	21	T Thr	4	7	4	7
G Gly	14	11	9	16	W Trp	8	3	8	3
H His	3	5	4	4	Y Tyr	8	5	9	4
I Ile	19	2	19	2	V Val	17	4	17	4

Zur Veranschaulichung wurde dieses Ergebnis von beiden Innen/Außen-Kriterien in einem Proteinviewer visualisiert und in Abbildung 15 gegenübergestellt. Einige Oberflächenbereiche des Proteins sind nach dem Kriterium für *außen* markiert. Dies resultiert aus Residuen, die vorrangig im Inneren des Proteins vorkommen, die Oberfläche nur mit wenigen Atomen berühren und somit eine geringe Wechselwirkung zum umgebenden Medium aufweisen. Beide Einfärbungen zeigen eine hohe Ähnlichkeit.

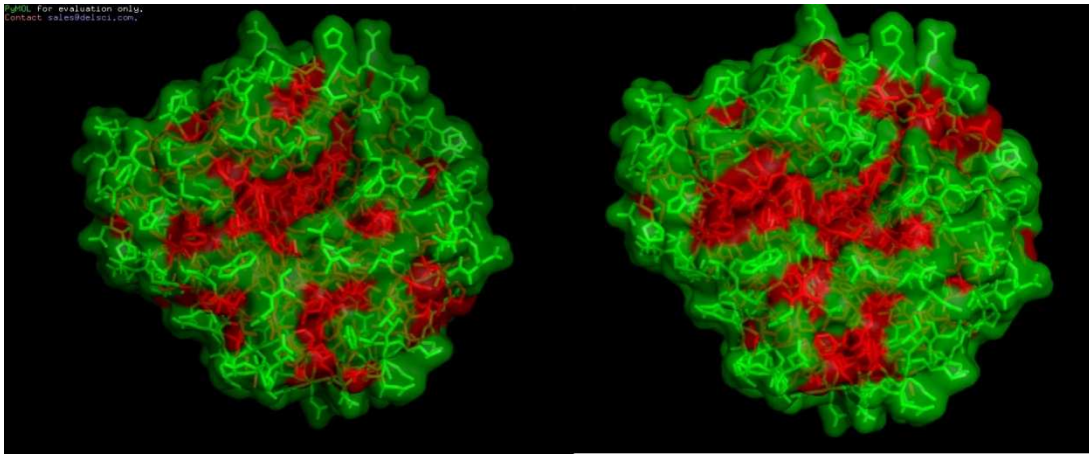


Abbildung 15: Vergleich beider Innen/Außen-Kriterien an einem Protein.

Alle Aminosäuren sind nach ihrer Orientierung eingefärbt, die durch die beiden Innen/Außen-Kriterien bestimmt wurden. Links ist das Protein (PDB-ID 1a3h) mit der Einfärbung dargestellt, die mit dem neu entwickelten Kriterium erzielt wurde. Rechts ist die Einfärbung des Proteins abgebildet, die mit dem ursprünglichen Kriterium entwickelt wurde. Grün sind alle Residuen, die „außen“ eingeordnet und rot alle Residuen, die „innen“ eingeordnet worden. Zusätzlich wurde die Oberfläche dargestellt und ebenfalls, entsprechend der Residue an der Oberfläche, eingefärbt.

4.3 Präferenzen zu Aminosäuren von Membranproteinen im Datensatz

Für den Vergleich zwischen Proteinbereichen, die außerhalb der Membran bzw. innerhalb der Membran in Membranproteinen vorkommen sowie für eine Gegenüberstellung von Membranproteinen und globulären Proteinen, wurden Präferenzen zu jeder kanonischen Aminosäure für die Sekundärstrukturelemente Helix und Faltblatt berechnet. Für diese Berechnungen ist eine Statistik zu dem gesamten Datensatz angefertigt worden. Diese Statistik wurde in Tabelle 7 visualisiert.

Die nahezu 60000 Residuen der Membranproteine im Datensatz sind nicht gleichmäßig verteilt. Obwohl mit 38,8 % über ein Drittel aller Residuen innerhalb der Membran liegen, ordnen sich dort nur 4,4 % in Loopbereichen und 4,5 % in Faltblättern an. Demnach bevorzugt die Mehrzahl aller Aminosäuren Helixstrukturen innerhalb der Membran.

Jede zweite Aminosäure, die sich außerhalb der Membran befindet, kommt in Loopstrukturen vor. Mit ca. 40 % sind etwas weniger aller Residuen außerhalb der Membran in Helices organisiert. Wie auch bei allen innerhalb der Membran vorkommenden Aminosäuren, ordnen sich auch bei allen außerhalb der Membran vorkommenden Residuen relativ wenig in Faltblättern an.

Insgesamt befinden sich nur 10,7 % aller Aminosäuren, die im Datensatz vorliegen, in Faltblättern. Aminosäuren in Loopregionen liegen zu 87,4 % außerhalb der Membran vor.

Tabelle 7: Verteilung der Aminosäuren im Datensatz.

Die Statistik differenziert zwischen den Sekundärstrukturelementen, in denen die Aminosäuren vorkommen und der Position der Aminosäuren - außerhalb oder innerhalb der Membran.

	außerhalb der Membran				innerhalb der Membran			
	Coil	Helix	Sheet	total	Coil	Helix	Sheet	total
Ala	1274	1399	200	2873	247	1916	250	2413
Arg	952	785	184	1921	90	274	119	483
Asn	983	576	125	1684	89	372	105	566
Asp	1080	619	209	1908	77	191	71	339
Cys	186	149	44	379	39	286	2	327
Gln	623	594	118	1335	53	254	93	400
Glu	1060	782	167	2009	54	199	85	338
Gly	1736	861	176	2773	281	1398	318	1997
His	362	290	62	714	33	224	41	298
Ile	786	830	294	1910	218	1944	110	2272
Leu	1325	1465	287	3077	281	2826	290	3397
Lys	1028	797	216	2041	45	207	89	341
Met	370	360	76	806	74	589	44	707
Phe	731	711	202	1644	184	1477	135	1796
Pro	1291	483	85	1859	124	440	31	595
Ser	1269	842	229	2340	172	931	199	1302
Thr	1070	821	257	2148	152	885	158	1195
Trp	253	269	69	591	53	453	60	566
Tyr	536	545	193	1274	109	566	205	880
Val	973	922	411	2270	209	1940	198	2347
gesamt	17852	14100	3604	35556	2584	17372	2603	22559
gesamte Anzahl Aminosäuren im Datensatz					58115			

Die Präferenzen wurden mit nachstehender Formel (6) ermittelt und sind in Tabelle 8 angegeben.

$$P_{rs} = \frac{\frac{n_{rs}}{n_s}}{\frac{n_r}{N}} \quad (6)$$

$P_{rs} \triangleq$ Präferenz einer Aminosäure r in einem Zustand s

$n_{rs} \triangleq$ Anzahl einer Aminosäure r im Zustand s

$n_s \triangleq$ Anzahl aller Aminosäuren im Zustand s

$n_r \triangleq$ Anzahl einer Aminosäure r im Datensatz

$N \triangleq$ Anzahl aller Aminosäuren im Datensatz

Tabelle 8: Präferenzen der kanonischen Aminosäuren in den Sekundärstrukturelementen. Die Angaben der Präferenzen der Aminosäuren in Membranproteinen sind in innerhalb der Membran liegender Aminosäuren (unter „in“) und in außerhalb der Membran liegender Aminosäuren (unter „out“) separiert. Die Präferenzen für gesamte Membranproteine sind unter „mp“ dargestellt. Zusätzlich sind die Präferenzen der Aminosäuren für globuläre Proteine unter „gp“ aufgetragen.

Aminosäure	Helix				Faltblatt			
	out	in	mp	gp	out	in	mp	gp
Alanin	1,23	1,03	1,16	1,45	0,69	0,90	0,80	0,97
Arginin	1,03	0,74	0,81	0,79	0,94	2,14	1,18	0,90
Asparagin	0,86	0,85	0,78	0,73	0,73	1,61	0,96	0,65
Asparaginsäure	0,82	0,73	0,67	0,98	1,08	1,82	1,17	0,80
Cystein	0,99	1,14	1,14	0,77	1,15	0,05	0,61	1,30
Glutamin	1,12	0,82	0,90	1,17	0,87	2,01	1,14	1,23
Glutaminsäure	0,98	0,76	0,77	1,53	0,82	2,18	1,01	0,26
Glycin	0,78	0,91	0,87	0,53	0,63	1,38	0,97	0,81
Histidin	1,02	0,98	0,94	1,24	0,86	1,19	0,95	0,71
Isoleucin	1,10	1,11	1,22	1,00	1,52	0,42	0,90	1,60
Leucin	1,20	1,08	1,22	1,34	0,92	0,74	0,83	1,22
Lysin	0,98	0,79	0,78	1,07	1,04	2,26	1,20	0,74
Methionin	1,13	1,08	1,16	1,20	0,93	0,54	0,74	1,67
Phenylalanin	1,09	1,07	1,17	1,12	1,21	0,65	0,92	1,28
Prolin	0,66	0,96	0,69	0,59	0,45	0,45	0,44	0,62
Serin	0,91	0,93	0,90	0,79	0,97	1,32	1,10	0,72
Threonin	0,96	0,96	0,94	0,82	1,18	1,15	1,16	1,20
Tryptophan	1,15	1,04	1,15	1,14	1,15	0,92	1,04	1,19
Tyrosin	1,08	0,84	0,95	0,61	1,49	2,02	1,73	1,29
Valin	1,02	1,07	1,14	1,14	1,79	0,73	1,23	1,65

4.3.1 Vergleich Präferenzen globulärer Proteine und Membranproteine

Zunächst wird der Vergleich zwischen Membranproteinen und globulären Proteinen betrachtet. Die Präferenzen zeigen deutlich, dass das Auftreten jeder Aminosäure in jedem Sekundärstrukturelement zwischen diesen Proteintypen variiert. Es ist zusätzlich erkennbar, dass diese Abweichungen zwischen den Präferenzen der Helices geringer ausfallen. Die durchschnittliche Abweichung in den Präferenzen beträgt hierbei 0,21. Faltblätter weisen mit 0,38 eine fast doppelt so hohe durchschnittliche Abweichung in den Präferenzen auf.

Somit ist anzunehmen, dass Energieprofile von Membranproteinen von den Energieprofilen von globulären Proteinen abweichen.

4.3.2 Vergleich Präferenzen innerhalb und außerhalb der Membran

In diesem Abschnitt werden ausschließlich Präferenzen von Aminosäuren, die sich außerhalb der Membran oder innerhalb der Membran befinden, verglichen. Bei der Betrachtung dieser Präferenzen fällt auf, dass, je nachdem, ob sich eine Residue außerhalb der Membran oder innerhalb der Membran befindet, sie unterschiedlich häufig in den Sekundärstrukturelementen vorkommt. Zusätzlich weichen die Präferenzen aller Aminosäuren in Helixstrukturen weniger stark voneinander ab, als die Präferenzen aller Aminosäuren in Faltblättern.

Präferenzen in Helices haben eine durchschnittliche Abweichung von 0,127 mit einer Standardabweichung von 0,1. Chou und Fasman haben Aminosäuren nach den Präferenzen in Gruppen eingeteilt (u. a. starke Helixbildner, schwache Helixbildner, indifferent, starke Helixbrecher) [12]. Durch die relativ geringe Abweichung ist sichergestellt, dass Aminosäuren sowohl außerhalb, als auch innerhalb der Membran die gleiche Tendenz aufweisen. Keine Aminosäure ist somit innerhalb der Membran ein Helixbildner und außerhalb der Membran ein Helixbrecher bzw. umgekehrt.

Eine wesentlich höhere durchschnittliche Abweichung mit 0,629 und einer Standardabweichung von 0,4 weisen die Aminosäuren in Faltblättern auf. Somit kann Glycin beispielsweise innerhalb der Membran als Faltblattbildner eingeteilt werden, außerhalb der Membran hingegen als Faltblattbrecher. Diese stark unterschiedliche Tendenz im Auftreten in Helices ist bei noch einigen anderen Aminosäuren, wie Arginin, Asparagin und Isoleucin erkennbar. Eine Ausnahme ist Prolin, das zwei Identische Werte aufweist. Diese Aminosäure kommt, sowohl innerhalb der Membran, als auch außerhalb der Membran, wenig in Faltblättern vor und wird somit als Faltblattbrecher eingeordnet.

Dieses unterschiedliche Vorkommen aller Aminosäuren in den Sekundärstrukturen unterstreicht die in der Energiefunktion unterschiedliche Handhabung von Residuen außerhalb bzw. innerhalb der Membran.

4.4 Vergleich des Berechnungsalgorithmus mit experimentellen Daten

In nachgestellter Abbildung 16 ist das Energieprofil von *Bacteriorhodopsin* visualisiert. Es veranschaulicht den Zusammenhang zwischen der Lösungsmittelzugänglichkeit und dem Energieprofilverlauf. Hydrophobe Abschnitte kommen in einem Protein häufig innerhalb der Membran vor, da es energetisch günstiger ist. Somit liegt die Hydrophobizität in Membranbereichen meist oberhalb des Durchschnittes. In dem Energieprofil ist ein gegenläufiges Verhalten zu dem Hydrophobizitätsplot zu erkennen. Grund hierfür ist, dass die freie Energie in Membranbereichen deutlich unterhalb des Durchschnittes liegt. Dieser Sachverhalt wird in Punkt 5.3 näher betrachtet. Im Hydrophobizitätsplot ist jedoch der Übergang von der vorletzten Helix in die letzte Helix nicht erkennbar. Im Energieprofil ist dieser Übergang deutlich zu sehen. Weiterhin sind alle Helices in kleinen Plateaus angeordnet, wobei Helices innerhalb der Membran einen niedrigeren Energiewert besitzen, als Helices außerhalb der Membran.

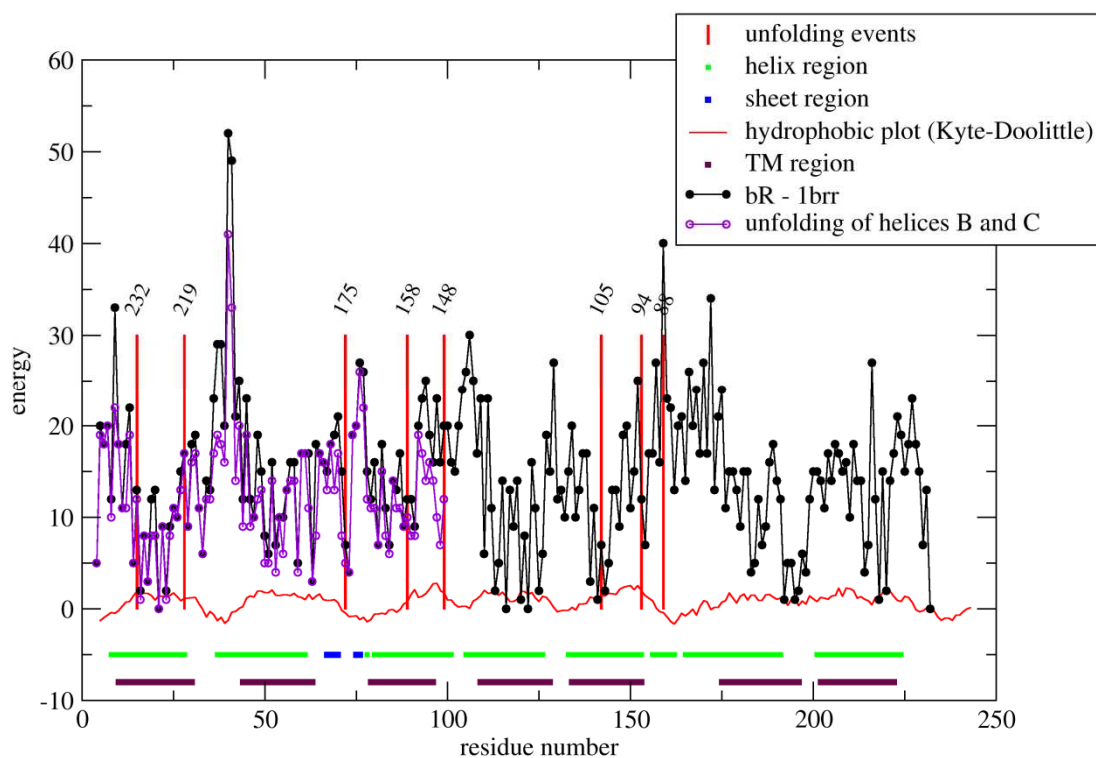


Abbildung 16: Energieprofil von *Bacteriorhodopsin* (PDB-ID: 1brr).

Das Energieprofil der Kette A von *Bacteriorhodopsin* (schwarz dargestellt), die Positionen der unfolding events, die Sekundärstrukturelemente, die Membranregion („TM region“) sowie der Hydrophobizitätsplot nach Kyle-Doolittle sind dargestellt.

5 Ergebnisse und Analyse der Energiefunktion

5.1 Allgemeine energetische Analyse von Membranproteinen

Die Berechnung der freien Energie wurde zu allen Proteinen im Datensatz nach dem in Punkt 3 beschriebenen Algorithmus durchgeführt. Aufgrund unvollständiger PDB-Dateien konnten aus 58115 im Datensatz vorhandenen Aminosäuren nur von 53389 Aminosäuren ein Energiewert berechnet werden. Die energetische Verteilung dieser Aminosäuren ist in Abbildung 17 dargestellt.

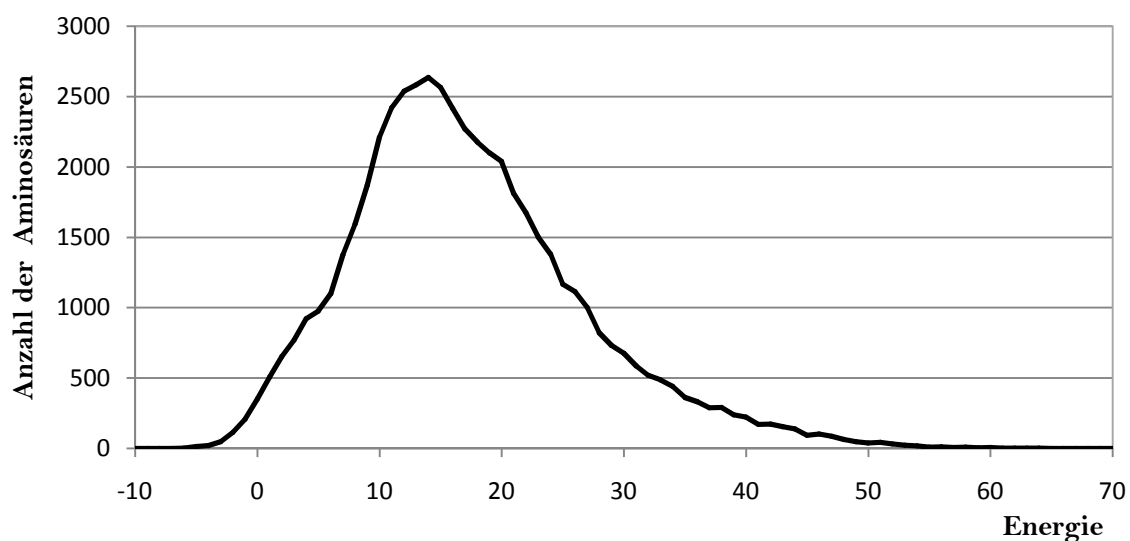


Abbildung 17: Energetische Verteilung aller Aminosäuren im Datensatz.

Der geringste Energiewert, der bei Glycin ermittelt wurde, beträgt -7,3856. Mit 65,3461 wurde der höchste Energiewert bei Lysin bestimmt. Durchschnittlich beträgt die Energie einer Aminosäure 17,2892 und weist eine Standardabweichung von 9,5585 auf.

Um eine bessere Klassifizierung und Einschätzung zu ermöglichen, wurden vier Bereiche, sogenannte Energie-Quantile, erstellt. Begrenzt werden sie durch das erste und dritte Quantil sowie dem Median aller Energiewerte des Datensatzes. Daraus ergeben sich folgende Bereiche:

- | | |
|----------------------------|----------------------------|
| 1. Energie-Quantil: | $x \geq 22,5473$ |
| 2. Energie-Quantil: | $22,5473 > x \geq 15,9948$ |
| 3. Energie-Quantil: | $15,9948 > x \geq 10,7655$ |
| 4. Energie-Quantil: | $10,7655 \geq x$ |

Die Einteilung von Aminosäuren in diese vier Quantile ermöglicht es, Rückschlüsse auf die Stabilität einer Aminosäure zu ziehen. Residuen mit einem Energiewert aus dem 4. Energie-Quantil weisen eine hohe Beständigkeit gegen physikochemische Eigenschaften auf. Aminosäuren hingegen, die Energiewerte aus dem 1. Energie-Quantil aufweisen, sind sehr instabil und an diesen Stellen kann die Polypeptidkette eher zu Veränderungen gebracht werden. Aus mittleren Energiewerten resultiert ein ambivalentes Verhalten der Aminosäuren.

Neben der allgemeinen Verteilung der Energiewerte der Aminosäuren wird nun zusätzlich der Energieprofilverlauf sowie die Lösungsmittelzugänglichkeit der Aminosäuren betrachtet. Dies ist in Abbildung 18 visualisiert.

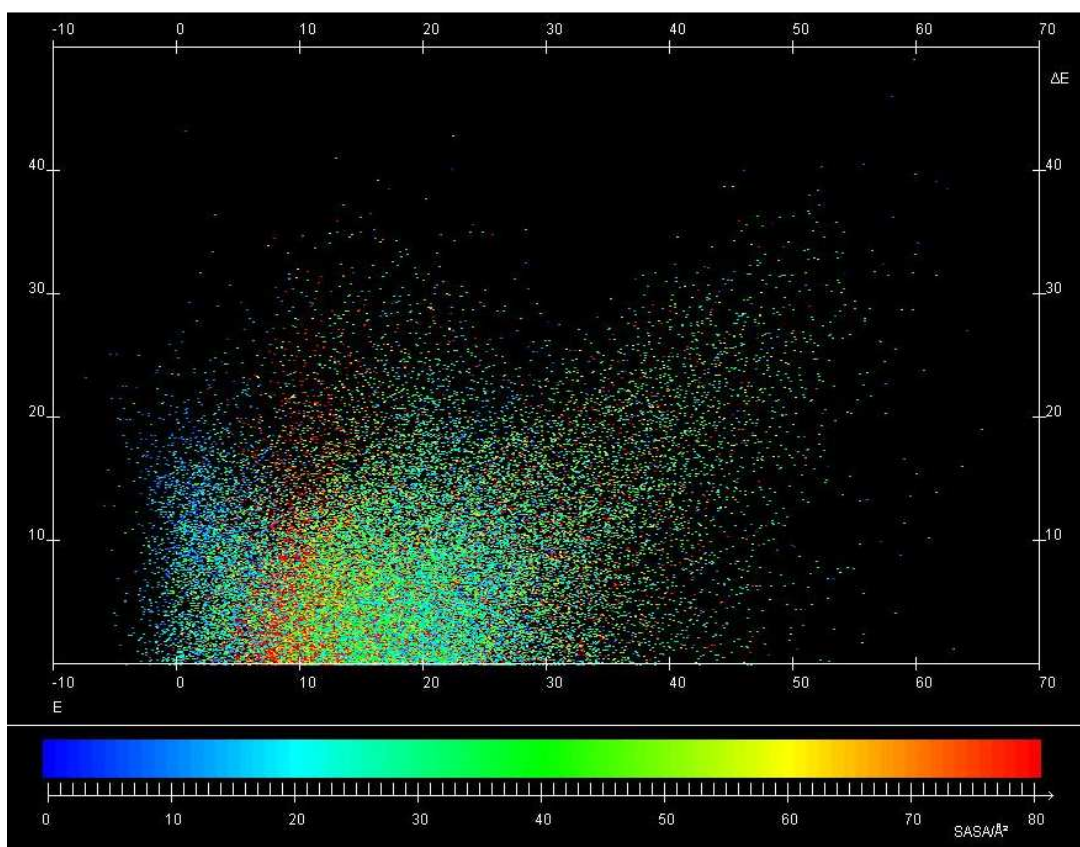


Abbildung 18: *E-ΔE-SASA Plot aller Aminosäuren.*

In diesem Plot ist der Energiewert einer Aminosäure i auf der Abszisse und die Differenz der Energiewerte der Aminosäure i und ihrer nachfolgenden Aminosäure $i+1$ auf der Ordinate aufgetragen. Es stellt einen Zusammenhang zwischen Energie und Energieprofilverlauf dar. Zusätzlich ist die SASA/Å² der Aminosäure i aufgetragen, die durch den POPS-Algorithmus berechnet wurde. Niedrige SASA/Å²-Werte werden blau, mittlere werden grün und hohe werden rot dargestellt. Übersteigt der Betrag 80 SASA/Å², bleiben die Werte rot.

Wie schon in Abbildung 17 ersichtlich ist, liegt ein Großteil aller Energiewerte im Bereich zwischen 5 und 25. Zusätzlich werden die Energieänderungen einer Aminosäure i

und ihrer nachfolgenden Aminosäure $i+1$ aufgezeigt. Energiedifferenzen bis zu 20 zwischen Aminosäure i und $i+1$ treten somit relativ häufig auf. Vereinzelt wurden sogar Differenzen von 30 oder mehr bestimmt.

Das gehäufte Auftreten der Datenpunkte aus Abbildung 18 in Bereichen mit niedrigen ΔE zeigt auf, dass plateauähnliche Verläufe in den Energieprofilen vorkommen. Diese Plateaus treten vorrangig in niedrigeren Energiebereichen, zwischen 0 und 25, auf.

Die farbliche Verteilung in Abbildung 18 lässt darauf schließen, dass bei Membranproteinen als Ganzes betrachtet, keine Aussage über den Zusammenhang zwischen der *SASA* einer Aminosäure und der Energie einer Aminosäure getroffen werden kann.

5.2 Energetische Analyse von Membran- und Nichtmembranbereichen

In Membranproteinen können Aminosäuren sowohl außerhalb der Membran, als auch innerhalb der Membran vorkommen. Die in Punkt 4.3.2 genannten Präferenzen zeigen die Möglichkeit auf, dass sich viele Aminosäuren strukturell unterschiedlich verhalten, je nachdem, ob sie innerhalb oder außerhalb der Membran lokalisiert sind. Ein kurze Aminosäuresequenz etwa, die innerhalb der Membran als Helix vorliegt, könnte außerhalb der Membran als Coilstruktur auftreten. Die Verteilung der Aminosäuren, relativ zur Membran, ist in Abbildung 17 dargestellt und wird in folgendem Abschnitt energetisch charakterisiert.

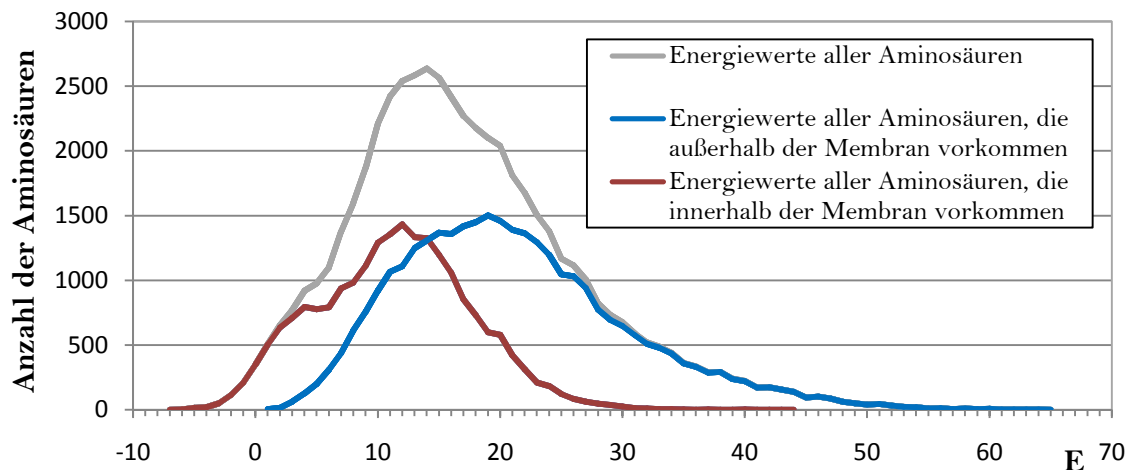


Abbildung 19: Energetische Verteilung äußerer bzw. innerer Aminosäuren.

Die Grafik zeigt die Verteilung der Energiewerte aller Aminosäuren, die sich außerhalb sowie innerhalb der Membran befinden. Zusätzlich ist die Verteilung der Energiewerte aller Residuen dargestellt. Die durchschnittliche Energie aller Residuen, die innerhalb der Membran auftreten beträgt 11,373 und weist eine Standardabweichung von 6,2944 auf. Alle außerhalb der Membran auftretenden Residuen weisen eine durchschnittliche Energie von 21,2186 und eine Standardabweichung von 9,33 auf.

Aus Abbildung 19 geht deutlich hervor, dass Aminosäuren, die sich innerhalb der Membran aufhalten, deutlich geringere Energiewerte aufweisen. Niedrige Energien deuten, wie in Punkt 5.1 bereits erwähnt, auf eine hohe Beständigkeit gegen physikochemische Eigenschaften hin. Aus den im Mittel höheren Energiewerten für außerhalb der Membran vorkommende Aminosäuren resultiert instabileres Verhalten.

Für eine weitere Analyse von beiden Gruppen⁵ wurde jeweils ein *E-ΔE-SASA* Plot angefertigt, der einen Zusammenhang zwischen den Energiewerten sowie dem Energieprofilverlauf und der Lösungsmittelzugänglichkeit darstellt. Die Plots sind in Abbildung 20 und Abbildung 21 aufgetragen. Zunächst wird auf die Auswertung innerhalb der Membran stehender Aminosäuren eingegangen, die in Abbildung 20 dargestellt sind.

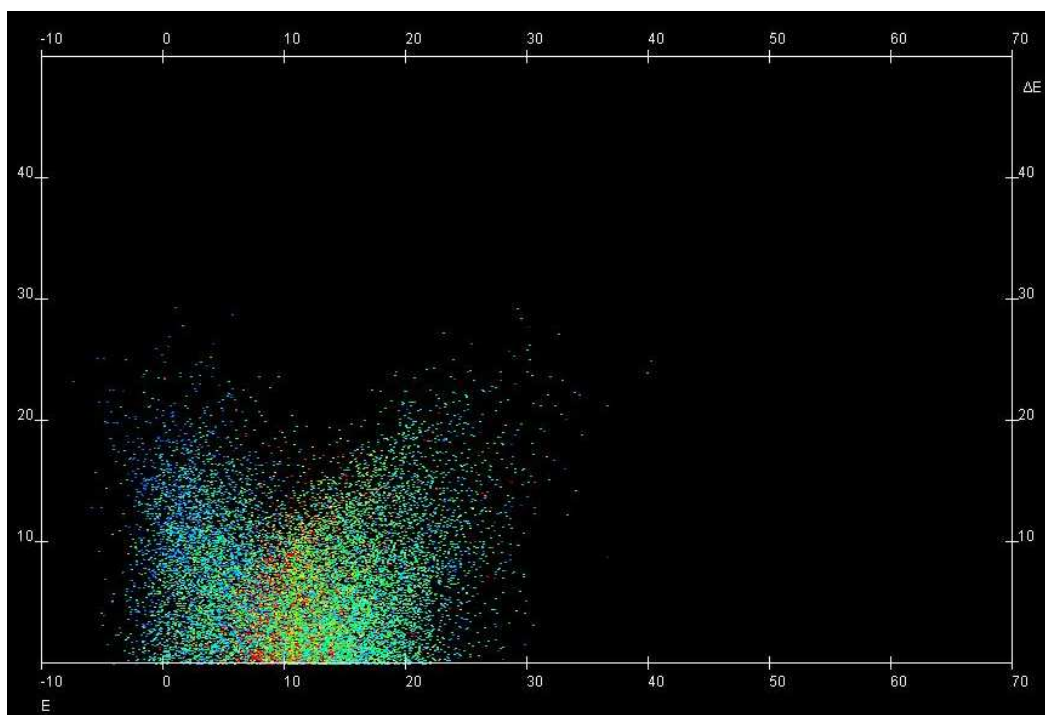


Abbildung 20: *E-ΔE-SASA* Plot aller Aminosäuren, die innerhalb der Membran liegen.

Aus Abbildung 20 ist ersichtlich, dass in der Membran liegende Aminosäuren eine relativ geringe energetische Streuung aufweisen. 68 % aller Energiewerte solcher Aminosäuren liegen zwischen 5,0786 und 17,6674. Die Energiedifferenzen sind im Vergleich zu den Energiedifferenzen aller Aminosäuren, die in Abbildung 18 dargestellt sind, relativ gering.

Aminosäuren, die innerhalb der Membran vorkommen, sind somit beständig bei niedrigen Energiewerten angesiedelt. Das Energieprofil in Abbildung 22 verdeutlicht diesen Sachverhalt.

⁵ Die Gruppe aller Aminosäuren, die außerhalb der Membran sowie die Gruppe aller Aminosäuren, die innerhalb der Membran liegen.

Abbildung 20 zeigt weiterhin, dass im gesamten Bereich sehr niedrige und mittlere Lösungsmittelzugänglichkeiten auftreten. Lediglich sehr hohe SASA/Å²-Werte treten vermehrt bei einer Energie um 10 auf. Eine Korrelation der Lösungsmittelzugänglichkeit und der Energiewerte ist somit auf den ersten Blick nicht zu erkennen. An dieser Stelle bedarf es weiterer Analysen, die nicht Bestandteil dieser Arbeit sind.

Die nachstehende Abbildung 21 beinhaltet, wie bereits erwähnt, einen *E-ΔE-SASA* Plot von außerhalb der Membran vorkommenden Aminosäuren.

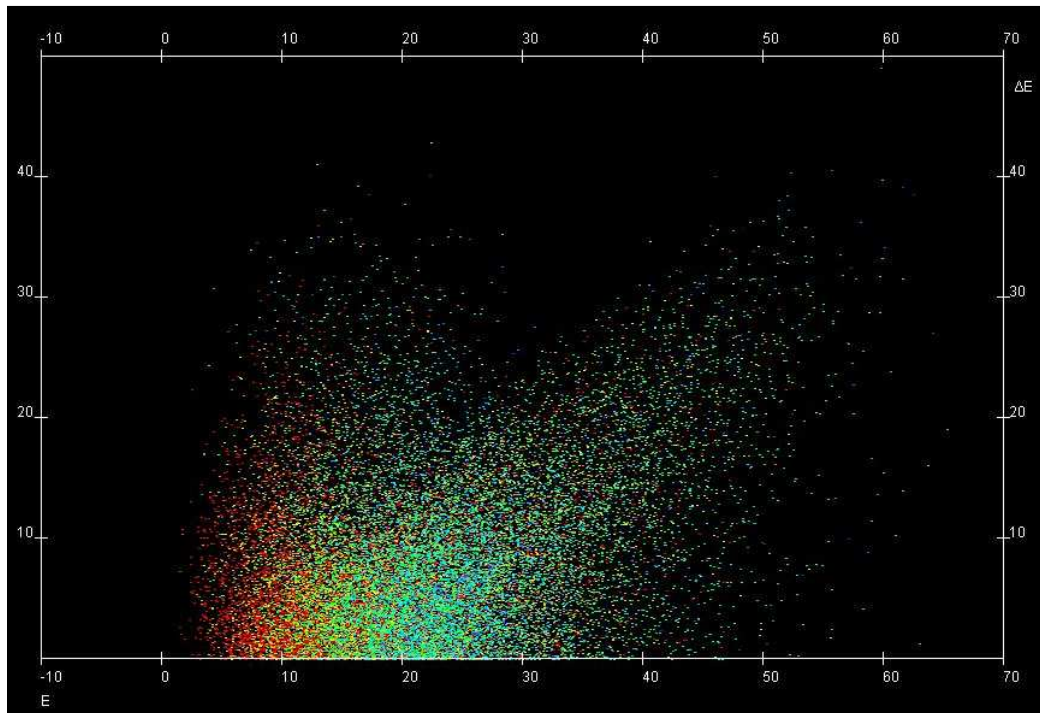


Abbildung 21: *E-ΔE-SASA* Plot aller Aminosäuren, die außerhalb der Membran liegen.

Aus Abbildung 21 geht hervor, dass Aminosäuren außerhalb der Membran eine relativ große energetische Streuung aufweisen. Von ihnen kommen 68 % zwischen einem Energiewert von 11,8886 und 30,5486 vor. Auch die Energiedifferenzen einer Aminosäure und ihrer nachfolgenden Aminosäure sind im Mittel größer als die Energiedifferenzen von transmembranen Aminosäuren. Abschnitte eines transmembranen Proteins, die außerhalb der Membran vorkommen, weisen somit höhere Energiewerte und größere Energiesprünge auf, als Bereiche, die innerhalb der Membran vorliegen. Das Energieprofil in Abbildung 22 unterstreicht dieses Resultat.

Weiterhin ist aus Abbildung 21 ersichtlich, dass mit zunehmenden Energiewerten die Lösungsmittelzugänglichkeit der Aminosäuren abnehmen.

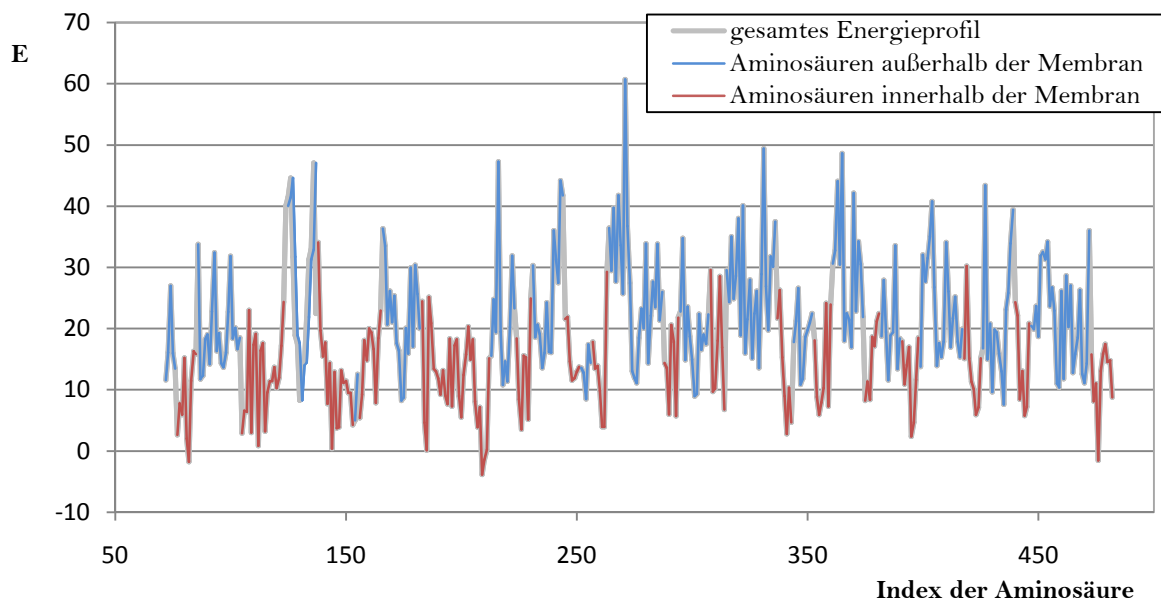


Abbildung 22: Energieprofil eines Proteins mit der PDB-ID 1a0s.

Bereiche, in denen die Aminosäuren außerhalb der Membran vorliegen, sind blau eingefärbt.

Bereiche, in denen die Aminosäuren innerhalb der Membran lokalisiert sind, sind rot markiert.

Übergangsbereiche sind schattiert.

Ein abschließender Vergleich innerhalb der Membran und außerhalb der Membran liegender Sequenzabschnitte von Proteinen verdeutlicht, dass innere Bereiche eine geringere freie Energie aufweisen und somit gegen äußere Einflüsse eine höhere Stabilität aufweisen, als Bereiche, die außerhalb der Membran vorkommen.

5.3 Energetische Analyse der Sekundärstrukturelemente

Die Verteilung der Sekundärstrukturelemente in den Proteinen im Datensatz wurde bereits unter Punkt 4.3 näher betrachtet. Für eine energetische Charakterisierung der Sekundärstrukturen ist eine Statistik zur Verteilung der Energiewerte in Tabelle 9 aufgetragen. Anhand dieser Werte ist kein signifikanter Unterschied zwischen Helixstrukturen, Coilstrukturen und Faltblättern zu erkennen.

Tabelle 9: Statistik zur energetischen Verteilung der Sekundärstruktur.

	Helix	Sheet	Coil
niedrigster Wert	-7,3856	-4,7628	-5,078
1. Quantil	10,2764	11,7826	11,2016
Median	15,2401	17,6968	16,7591
3. Quantil	21,5026	24,4142	23,4077
höchster Wert	64,1535	61,6595	65,3461
Durchschnitt	16,5108	18,6697	18,0263
Standardabweichung	-9,3072	-10,3245	-9,4276
Anzahl	29733	8864	14792

Für eine genauere Betrachtung wurde in einer weiteren Statistik zusätzlich unterschieden, ob die Sekundärstrukturen innerhalb oder außerhalb der Membran vorkommen. In Tabelle 10 ist diese erweiterte Statistik visualisiert. Es ist auch bei den Sekundärstrukturelementen zu erkennen, dass Bereiche, die innerhalb der Membran vorkommen, wesentlich geringere freie Energie vorweisen, als jene, die außerhalb der Membran vorkommen. Werden alle Sekundärstrukturelemente innerhalb der Membran bzw. außerhalb der Membran für sich betrachtet, ist allerdings auch bei der erweiterten Statistik, kein signifikanter Unterschied zu erkennen.

Tabelle 10: Statistik zur energetischen Verteilung der Sekundärstruktur relativ zur Membran. Die Sekundärstrukturelemente sind zwischen Membran-inneren und Membran-äußeren differenziert.

	außerhalb der Membran			innerhalb der Membran		
	Helix	Sheet	Coil	Helix	Sheet	Coil
niedrigster Wert	1,07	2,4211	0,8918	-7,3856	-4,7628	-5,0779
erstes Quantil	15,0569	17,1711	12,6816	7,2446	6,4515	5,3749
Median	20,6019	22,3659	18,2418	11,6761	11,6701	9,8717
drittes Quantil	26,9539	29,0844	24,7436	15,6746	15,7973	14,1052
höchster Wert	64,1535	61,6595	65,3461	43,7533	39,468	36,6859
Durchschnitt	21,8124	23,7471	19,486	11,5666	11,4268	9,9749
Standardabweichung	-9,1789	-9,2917	-9,1922	-6,179	-6,8023	-6,0412
Anzahl	14348	5211	12522	15385	3653	2270

Infolgedessen ist es nicht möglich, ein Energieprofil nur nach dem Verlauf in Sekundärstrukturen einzuteilen.

5.4 Energetische Charakterisierung von Aminosäuren

5.4.1 Anlegen energetischer Präferenzen von Aminosäuren

Für eine Analyse wurden die Energiewerte jeder Aminosäure im gesamten Datensatz untersucht. Dabei wurde unterschieden, wo sich die Aminosäuren relativ zur Membran und in welchem Sekundärstrukturelement sie sich befanden. Für eine einfachere Einteilung der Aminosäuren entsprechend ihrer freien Energie wurden die Energie-Quantile, die unter Punkt 5.1 vorgestellt wurden, genutzt. Daraus resultieren zwölf Zustände für Aminosäuren innerhalb der Membran und zwölf Zustände für Aminosäuren außerhalb der Membran. Diese Einteilung ermöglicht Rückschlüsse auf die Stabilität und Charakterisierung einer Aminosäure. In Tabelle 11 und Tabelle 12 sind diese Zustände⁶ aufgeführt.

Tabelle 11: Präferenzen der Energiewerte von Aminosäuren innerhalb der Membran.

Die Präferenzen differenzieren zwischen den Sekundärstrukturelementen und zusätzlich zwischen den Energie-Quantilen. Letztere sind jeweils in aufsteigender Reihenfolge von links nach rechts angeordnet und mit EQ gekennzeichnet. Die höchsten Werte der Präferenzen jedes Sekundärstrukturelementes sind farblich markiert. In dieser Statistik sind nur Aminosäuren, die innerhalb der Membran vorkommen, vertreten.

	Helixstrukturen				Faltblattstrukturen				Loopstrukturen			
	1. EQ	2. EQ	3. EQ	4. EQ	1. EQ	2. EQ	3. EQ	4. EQ	1. EQ	2. EQ	3. EQ	4. EQ
Gly	0	0,011	0	2,056	0,124	0,016	0	2,91	0	0	0	2,094
Ala	0	0,009	0,104	2,253	0	0	0,104	2,059	0	0	0,129	1,578
Cys	0	0,023	0,266	2,442	0	0	0,06	0,393	0	0	0,108	2,068
Pro	0,199	0,062	0,582	1,741	0	0	0,353	1,071	0	0	1,154	2,714
Ser	0,062	0,023	0,014	2,092	0,095	0,025	0,045	2,611	0	0	0,027	2,116
Thr	0,067	0,149	1,077	1,329	0	0,192	1,387	1,334	0	0,429	0,928	1,409
Tyr	0,267	0,69	1,324	0,599	0,551	0,987	2,715	1,25	0,466	0,327	1,782	0,881
Val	0,272	1,193	1,576	0,736	0,473	0,879	1,102	0,566	0,178	0,936	1,242	0,512
Met	0,17	0,707	1,825	0,751	0,175	0,325	1,394	0,495	0,592	0,83	1,818	0,533
Asn	0,488	1,04	1,56	0,326	0,215	1,543	2,426	0,489	0	1,407	2,181	1,002
Glu	0,467	0,697	1,261	0,487	0,36	1,628	3,499	0,696	0	1,071	2,131	0,944
Phe	0,089	0,479	1,597	1,014	0	0,439	1,263	0,751	0	0,573	1,397	0,672
Asp	0,232	0,974	1,135	0,388	0	1,713	2,58	0,61	0	2,129	2,723	1,824
His	0	0,111	1,269	1,017	0	0,244	2,009	1,509	0	0,272	1,549	1,334
Trp	1,231	2,309	1,311	0,199	1,117	2,612	1,259	0,228	0,757	1,461	1,07	0,52
Ile	0,864	1,997	1,631	0,361	0,49	1,027	0,989	0,117	0,554	1,553	1,503	0,404
Gln	0,861	1,393	1,259	0,256	0,591	4,474	1,804	0,235	1,002	2,283	2,08	0,602
Arg	3,787	1,699	0,643	0,135	6,22	5,551	1,246	0	3,373	5,174	1,681	0,47
Leu	3,62	2,492	1,219	0,117	2,105	2,024	1,096	0,07	3,443	2,156	0,838	0,185
Lys	10,17	1,786	0,239	0,01	23,73	2,811	0,171	0,041	17,32	2,386	0,617	0,212

⁶ In den Zuständen sind die Präferenzen der Aminosäuren aufgetragen, die mit Formel (6) berechnet wurden.

Bei der Betrachtung von Tabelle 11 fällt die Verteilung der markierten Werte auf. Fast alle Aminosäuren besitzen die Maxima aller drei Sekundärstrukturen im gleichen Energie-Quantil. Lediglich Threonin und Arginin haben ihre Höchstwerte in je zwei verschiedenen Energie-Quantilen, wobei diese benachbart sind. Diese Verteilung der markierten Werte korreliert mit den Eigenschaften der Aminosäuren, die in Tabelle 1 sowie in dem Venn-Diagramm in Abbildung 1 aufgeführt sind.

Auch bei der Charakterisierung der Aminosäuren aus Tabelle 12 fällt auf, dass die Mehrzahl aller Aminosäuren ihre Maxima in allen Sekundärstrukturelementen im gleichen Energie-Quantil vorweisen können. Fünf Aminosäuren haben Höchstwerte in zwei benachbarten Energie-Quantilen vorzuweisen und lediglich Tryptophan und Prolin unterliegen einer breiteren Verteilung ihrer Höchstwerte.

Tabelle 12: Präferenzen der Energiewerte von Aminosäuren außerhalb der Membran.

Die Präferenzen differenzieren ebenfalls zwischen den Sekundärstrukturelementen und den Energie-Quantilen. Letztere sind auch hier in aufsteigender Reihenfolge von links nach rechts angeordnet und mit EQ gekennzeichnet. Die höchsten Werte der Präferenzen jedes Sekundärstrukturelementes sind farblich markiert. In dieser Statistik sind nur Aminosäuren, die außerhalb der Membran vorkommen, vertreten.

	Helixstrukturen				Faltblattstrukturen				Loopstrukturen			
	1. EQ	2. EQ	3. EQ	4. EQ	1. EQ	2. EQ	3. EQ	4. EQ	1. EQ	2. EQ	3. EQ	4. EQ
Gly	0,613	0,857	1,186	1,645	0,583	0,847	0,88	1,222	0,83	1,185	1,375	1,865
Pro	1,088	0,744	0,693	0,349	0,824	0,565	0,722	0,842	1,692	1,192	1,181	0,869
Ala	0,548	1,623	1,664	1,648	0,385	1,07	1,194	1,754	0,262	0,94	1,227	1,451
Cys	0	0,172	1,788	6,066	0	0,228	2,849	10,52	0	0,211	1,359	3,901
Ser	0,754	1,022	1,036	0,793	1,069	0,873	1,235	0,786	0,865	1,14	1,386	1,179
Thr	0,937	1,031	1,002	0,534	1,214	1,077	0,76	0,981	1,09	1,166	0,883	0,897
Tyr	0,771	1,263	1,182	0,856	1,28	1,844	1,001	0,769	0,59	0,991	0,909	0,935
Val	0,238	1,333	1,651	1,961	0,501	2,453	2,66	1,669	0,151	0,899	1,19	1,322
Met	1,417	1,25	0,707	0,599	1,129	0,959	0,495	0	0,874	1,018	0,841	0,582
Asn	1,38	0,743	0,424	0,189	1,263	0,469	0,157	0,249	1,857	1,082	0,948	0,52
Glu	1,72	0,693	0,372	0,187	1,363	0,519	0,199	0,211	1,723	0,86	0,76	0,441
Phe	0,169	1,166	1,995	2,758	0,229	1,439	2,825	2,582	0,122	0,868	1,202	1,905
Asp	1,172	0,824	0,537	0,337	1,38	0,789	0,325	0,222	1,404	1,221	1,056	0,743
His	1,594	0,868	0,395	0,191	1,683	0,377	0,252	0,401	1,202	0,948	0,908	0,735
Trp	0,546	1,236	1,802	1,415	0,588	1,986	1,482	1,241	0,263	1,047	1,103	1,203
Ile	0,321	1,369	1,756	1,889	0,498	2,167	2,25	1,681	0,201	0,973	1,178	1,097
Gln	2,107	0,534	0,251	0,098	1,473	0,195	0,13	0,311	1,792	0,761	0,596	0,269
Arg	1,99	0,458	0,116	0,069	1,846	0,17	0,023	0,144	1,916	0,846	0,528	0,299
Leu	0,895	1,395	1,27	1,138	0,794	1,226	1,091	0,777	0,645	1,054	0,812	0,972
Lys	1,957	0,439	0,128	0,065	2,008	0,181	0,128	0	2,042	0,811	0,438	0,129

5.4.2 Auswertung der energetischen Präferenzen

Ein Vergleich beider Gruppen⁵ ist in diesem Abschnitt gemäß der Einordnung der vier Energie-Quantile in *instabil*, *ambivalent* und *stabil* erfolgt. Für die Einordnung einer Aminosäure in Energie-Quantile wurden die drei Maxima der Präferenzen der Sekundärstrukturelemente herangezogen.

Grundsätzlich ist zu erkennen, dass kleine Aminosäuren innerhalb der Membran meist im 4. Energie-Quantil, im stabilen Bereich, vorzufinden sind. Die sehr kleinen Aminosäuren Glycin, Alanin, Cystein und Serin weisen ihre Höchstwerte sogar ausschließlich aus dem 4. Energie-Quantil auf. Begründen lässt sich dies damit, dass kleinere Aminosäuren vorrangig im Inneren von Proteinen vorkommen, um als „Lückenfüller“ eine kompakte Struktur der Proteine zu gewährleisten. Innerhalb dieser Proteine sind physikochemische Wechselwirkungen mit der Umgebung, durch die stabilisierende Wirkung benachbarter Aminosäuren, sehr gering. Zusätzlich ist die Größe von kleinen Aminosäuren an sich von Bedeutung. Großen Aminosäuren, wie beispielsweise Tryptophan, ist es aus dem Inneren von Proteinen möglich, bis an die Oberfläche von Proteinen zu reichen, an der sie Wechselwirkungen mit der Umgebung ausgesetzt sind. Kleineren Aminosäuren ist dies nicht möglich. Dieser Sachverhalt ist noch etwas stärker ausgeprägt, liegt eine Aminosäure außerhalb der Membran. Dort ist keine Doppellipidschicht vorhanden, die einzelne Aminosäuren zusätzlich stabilisieren kann. Aus diesem Grund sind nicht alle kleinen Aminosäuren, die außerhalb der Membran vorkommen, im energetisch stabilen oder ambivalenten Bereich und nicht alle vier sehr kleinen Aminosäuren sind im stabilen Bereich, vorzufinden.

Geladene Aminosäuren kommen innerhalb der Membran im instabilen bzw. ambivalenten Bereich vor. Durch die Membran liegt für die geladenen Aminosäuren auch hier eine geringe Lösungsmittelzugänglichkeit vor. Einzig durch ihre Ladung sind sie in instabilen und ambivalenten Energie-Quantilen einzuordnen. Geladene Aminosäuren außerhalb der Membran sind dagegen alle im instabilen Bereich vorzufinden. Dieser Sachverhalt resultiert aus höherer Lösungsmittelzugänglichkeit in Bereichen außerhalb der Membran. Durch ihre Ladung können diese Aminosäuren stärker als andere Aminosäuren mit ihrer Umgebung wechselwirken, und haben daher einen instabilen Charakter.

Bei polaren Aminosäuren sind ebenfalls Unterschiede zwischen beiden Gruppen zu verzeichnen. Diese Aminosäuren treten im Inneren der Membran im gesamten Bereich auf, sind vermehrt aber im ambivalenten Bereich, tendierend zum stabilen Bereich, anzutreffen. Grund hierfür ist wohl auch in diesem Fall die Membran, die vor Wechselwirkungen mit der Umgebung schützt. Gegenläufig verhält es sich mit polaren Aminosäuren außerhalb der

Membran, die ebenfalls im gesamten Bereich auftreten. Jedoch ist die Mehrzahl der polaren Aminosäuren im 1. Energie-Quantil anzutreffen.

Aminosäuren mit hydrophobem Charakter sind sowohl innerhalb der Membran als auch außerhalb der Membran im gesamten Energiebereich anzutreffen. Grund hierfür ist, dass andere Eigenschaften der Aminosäuren, wie Ladung und Polarität, bei Wechselwirkungen mit der Umgebung, überwiegen.

Abschließend lässt sich feststellen, dass Aminosäuren, die innerhalb der Membran vorkommen, stabiler sind, als Aminosäuren, die sich außerhalb der Membran befinden. Die Membran nimmt somit eine zentrale Stellung für die Stabilität von Membranproteinen ein.

5.4.3 Korrelationsmatrizen der Aminosäuren

Um alle Aminosäuren untereinander vergleichen zu können, wurden zwei Korrelationsmatrizen angefertigt. Dazu wurden alle Energiewerte jeder kanonischen Aminosäure in 81 gleich große Klassen⁷ eingeteilt. Anhand dieser 81 Klassen wurde der empirische Korrelationskoeffizient für die Energiewerte zwischen jedem kanonischen Aminosäurepaar, jeweils getrennt für die Aminosäuren außerhalb bzw. innerhalb der Membran, angefertigt. Somit ergeben sich zwei Matrizen, die in den Anlagen in Tabelle 14 und Tabelle 15 dargestellt sind. Die gesamte Statistik zu jeder einzelnen Aminosäure ist in Tabelle 13, ebenfalls in den Anlagen, dargestellt.

Anhand dieser Koeffizienten ist zu sehen, dass Aminosäuren, die ähnliche Eigenschaften aufweisen, teilweise in ihrem Energiewerten korrelieren. Das ist an einem Beispiel in Abbildung 23 dargestellt. Auch hier sind zwischen den Aminosäuren, die außerhalb bzw. innerhalb der Membran liegen, partiell große Unterschiede zu erkennen. Grundsätzlich korrelieren die Aminosäuren stärker miteinander, wenn sie außerhalb der Membran stehen. Der durchschnittliche Korrelationskoeffizient aller Paarungen außerhalb der Membran beträgt 0,7 und innerhalb der Membran 0,494. Außerhalb der Membran korreliert Valin durchschnittlich am stärksten mit den 19 anderen Aminosäuren. Es hat einen durchschnittlichen Koeffizienten von 0,671. Glycin korreliert außerhalb der Membran mit 0,013 durchschnittlich am schwächsten mit anderen Aminosäuren. Innerhalb der Membran korrelieren Prolin am stärksten und Cystein am schwächsten mit anderen Aminosäuren. Ihre durchschnittlichen Koeffizienten betragen 0,813 und 0,292.

⁷ Diese Klassen verliefen von $-10 \pm 0,5$ bis $70 \pm 0,5$. Die kleinste Klasse hatte somit bei $-10,5$ und $-9,5$ ihre Grenzen und die größte Klasse bei $69,5$ und $70,5$. Jeweils die untere Grenze gehörte zu der jeweiligen Klasse dazu (Bsp.: $18,5 \leq x < 19,5$).

Energiewerte zweier Aminosäuren, Asparaginsäure und Glutaminsäure, sind in nachstehender Abbildung 23 dargestellt. Beide Aminosäuren sind polar und negativ geladen. Einziger wesentlicher Unterschied ist, dass Asparaginsäure zu den kleineren Aminosäuren gehört. Es ist erkennbar, dass sowohl alle Energiewerte innerhalb der Membran als auch alle Energiewerte außerhalb der Membran, stark korrelieren. Der Korrelationskoeffizient von beiden Aminosäuren beträgt innerhalb der Membran 0,98 und außerhalb der Membran 0,93.

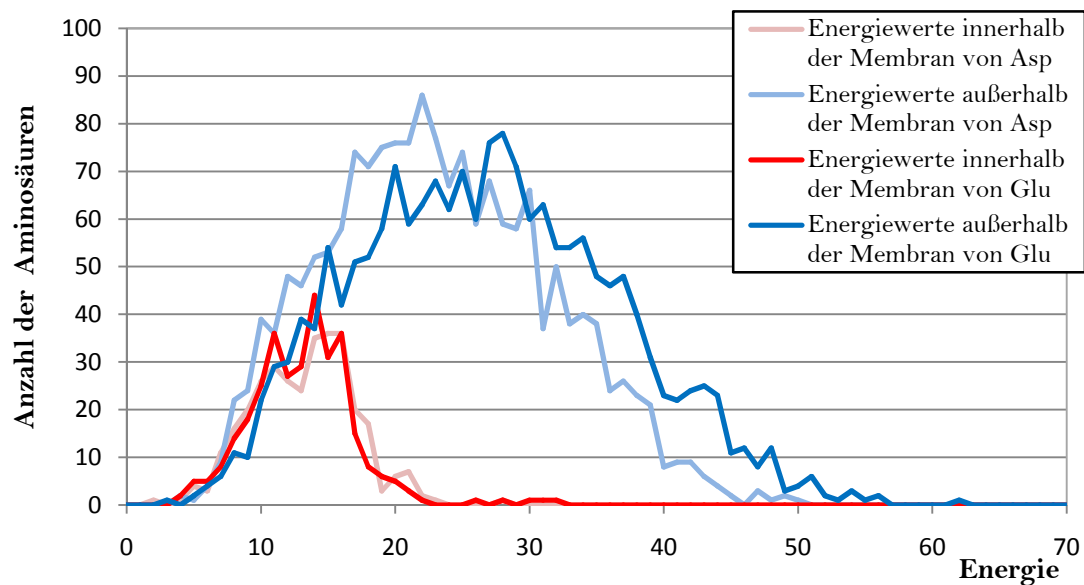


Abbildung 23: Vergleich der Energieverteilung von Asparaginsäure und Glutaminsäure.

Werden die Energiewerte zweier Aminosäuren verglichen, die nicht wesentliche Eigenschaften teilen, sind große Unterschiede zu erkennen. In nebenstehender Abbildung 24 sind Lysin und Glycin gegenübergestellt. Deutlich zu erkennen ist, dass die Energiehäufigkeiten innerhalb der Membran zwischen beiden Aminosäuren stark voneinander abweichen. Dort ist ein Korrelationskoeffizient von $-0,18$ festzustellen. Außerhalb der Membran sind die Verteilungen beider Aminosäuren in den Energiewerten nicht ganz so unterschiedlich, aber dennoch markant. Beide Aminosäuren weisen an dieser Stelle einen Korrelationskoeffizient von $0,26$ auf.

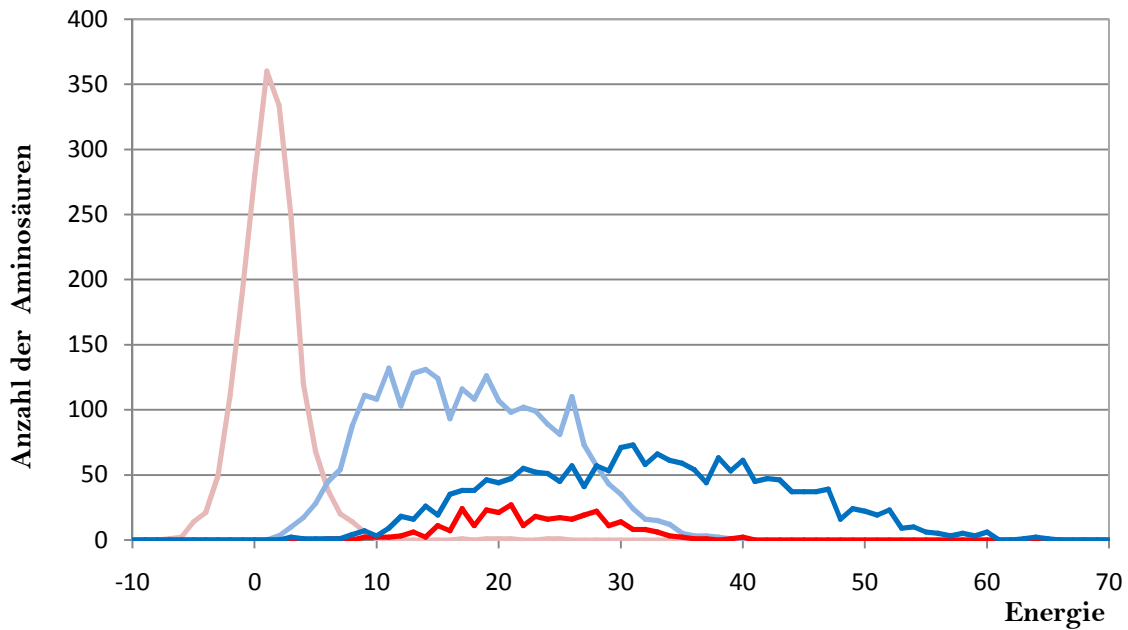


Abbildung 24: Vergleich der Energieverteilung von Glycin und Lysin.

Haben Aminosäuren energetisch eine ähnliche Verteilung in den Energiewerten, so ist anzunehmen, dass beide Aminosäuren ähnlich auf physiologische Umstände reagieren. Kommt es zu einer Mutation in einem Protein, so dürfte ein Austausch von Aminosäuren, die zu gleichen Energiewerten tendieren, unter Umständen keinen Funktionsverlust des Proteins bewirken. Weiterhin ist in Tabelle 14 und Tabelle 15 zu sehen, dass Aminosäuren innerhalb der Membran wesentlich signifikanter auftreten und somit weniger oft isofunktionell austauschbar sind.

6 Ausblick

Energieprofile besitzen ein hohes Maß an Struktur- und physikochemischen Informationen des Ausgangsproteins. Es ist nun zu prüfen, ob ein zuverlässiger Algorithmus zur Sekundärstrukturvorhersage auf Grundlage dieser Profile entwickelt werden kann.

Einen Schritt weiter geht der Gedanke, auf Basis einer Aminosäuresequenz, ein Energieprofil vorherzusagen und aus diesem Energieprofil Rauminformationen abzuleiten. Mit diesen gewonnenen Informationen könnten strukturell ähnliche Proteine, trotz geringer Sequenzidentitäten, anhand der Daten aus Energieprofilen von bekannten Proteinen, identifiziert werden. Die Grundlage für dieses Vorgehen würde der GOR-Algorithmus liefern.

Weiterhin ist es vorstellbar, dass Proteine durch energieprofilbasierende Analysen eindeutiger zu Proteinfamilien zugeordnet bzw. eine heuristische Suche auf Grundlage von Energieprofilen erstellt werden könnte.

Eine genaue Analyse des Zusammenhangs zwischen dem Energiewert und der Lösungsmittelzugänglichkeit bei Proteinbereichen, die außerhalb der Membran bzw. innerhalb der Membran liegen, sollte an diese Arbeit anschließen.

Anlagen

Tabelle 13: Gesamte Statistik der Energiewerte der Aminosäuren.

Für jede kanonische Aminosäure ist der niedrigste Energiewert, das 1. Quantil, der Median, das 3. Quantil, der höchste Energiewert, der durchschnittliche Energiewert sowie die Standardabweichung um den durchschnittlichen Energiewert angegeben. Diese Angaben sind zwischen allen Residuen, die innerhalb bzw. außerhalb der Membran vorkommen, unterteilt. Innerhalb dieser beiden Bereiche sind die Angaben zusätzlich in die Sekundärstrukturelemente Helix, Faltblatt und Schleife (in der Tabelle aus Platzgründen als „Helix“, „Sheet“ und „Coil“ vermerkt) separiert. Die Angaben für einen gesamten Bereich sind jeweils in der Spalte „gesamt“ aufgeführt. In der letzten Spalte sind die Angaben für die gesamte Residue angegeben.

	innerhalb der Membran				außerhalb der Membran				gesamte
	gesamt	Helix	Sheet	Coil	gesamt	Helix	Sheet	Coil	Residue
Ala Anzahl	2286	1695	369	222	2613	1387	334	892	4899
niedrigster Wert	-0,97	-0,97	0,37	0,51	1,07	1,07	4,70	2,40	-0,97
1. Quantil	4,17	4,33	3,61	3,77	12,28	13,32	14,06	10,14	5,69
Median	5,65	5,76	5,16	5,30	16,58	17,49	18,10	14,31	9,54
2. Quantil	7,19	7,20	7,17	6,91	20,72	21,49	22,36	18,89	17,06
höchster Wert	18,68	18,68	14,53	13,61	37,03	37,03	33,85	31,73	37,03
Durchschnittswert	5,85	5,96	5,53	5,54	16,56	17,42	18,15	14,63	11,56
Standardabweichung	2,40	2,34	2,61	2,46	5,83	5,55	5,65	5,84	7,03
Arg Anzahl	468	251	141	76	1764	782	277	705	2232
niedrigster Wert	3,58	4,93	10,94	3,58	2,77	2,84	8,65	2,77	2,77
1. Quantil	14,75	14,03	16,31	12,57	22,88	25,20	29,97	18,30	19,20
Median	17,89	17,64	18,68	16,20	30,38	31,16	36,24	25,68	26,65
2. Quantil	20,79	20,57	21,18	18,99	38,40	38,74	42,19	33,90	35,98
höchster Wert	31,65	31,65	28,55	27,88	63,01	63,01	61,66	62,53	63,01
Durchschnittswert	17,75	17,72	18,90	15,72	30,57	32,10	36,05	26,74	27,89
Standardabweichung	4,77	5,06	3,81	4,75	10,63	9,62	8,89	10,97	11,02
Asn Anzahl	541	342	120	79	1533	613	199	721	2074
niedrigster Wert	3,25	3,25	6,75	3,93	3,74	3,74	9,54	5,09	3,25
1. Quantil	11,43	11,89	11,70	9,56	18,52	19,97	22,87	15,84	14,54
Median	13,75	14,05	13,71	11,80	24,59	25,32	29,72	22,48	20,60
2. Quantil	15,88	16,04	15,85	14,32	31,66	32,11	34,73	29,84	28,96
höchster Wert	38,23	38,23	23,85	21,44	50,91	49,91	45,52	50,91	50,91
Durchschnittswert	13,75	14,10	13,87	12,04	25,00	25,91	28,89	23,15	22,06
Standardabweichung	3,68	3,75	3,12	3,77	8,76	8,08	7,89	9,07	9,20
Asp Anzahl	325	174	79	72	1722	668	274	780	2047
niedrigster Wert	2,48	5,01	7,67	2,48	4,05	6,70	8,85	4,05	2,48
1. Quantil	10,37	10,89	11,55	8,98	16,99	18,78	20,56	14,39	14,89
Median	13,52	13,73	13,90	11,33	22,67	23,73	26,63	19,98	20,73
2. Quantil	15,90	16,11	15,60	14,15	29,04	29,93	32,03	26,54	27,76
höchster Wert	28,12	28,12	21,94	20,91	49,99	49,99	44,84	48,93	49,99
Durchschnittswert	13,23	13,59	13,91	11,60	23,18	24,41	26,67	20,91	21,60
Standardabweichung	3,80	3,78	3,25	3,97	8,36	7,99	7,40	8,35	8,62

Cys Anzahl	303	255	10	38	351	151	58	142	654
niedrigster Wert	1,39	1,39	5,22	1,94	0,89	1,11	3,33	0,89	0,89
1. Quantil	5,61	5,70	5,80	4,38	6,91	7,24	9,38	6,10	6,17
Median	7,15	7,19	6,37	6,77	9,85	9,76	10,76	8,96	8,16
2. Quantil	8,42	8,42	7,19	8,16	12,44	12,85	13,23	11,58	10,78
höchster Wert	16,55	16,55	13,01	11,76	21,03	19,18	18,02	21,03	21,03
Durchschnittswert	7,19	7,29	7,35	6,46	9,83	10,06	11,19	9,05	8,61
Standardabweichung	2,40	2,39	2,36	2,40	3,91	3,82	3,13	4,14	3,55
Gln Anzahl	394	236	105	53	1229	600	163	466	1623
niedrigster Wert	4,08	5,88	8,01	4,08	5,52	6,43	5,67	5,52	4,08
1. Quantil	12,59	12,48	13,94	10,71	21,87	24,25	27,19	18,04	16,76
Median	14,93	14,79	16,26	13,05	29,17	30,31	32,81	24,90	24,75
2. Quantil	17,86	17,91	18,11	16,43	36,04	36,58	39,59	32,51	33,42
höchster Wert	43,75	43,75	23,92	22,91	56,43	52,95	56,43	52,06	56,43
Durchschnittswert	15,25	15,27	16,20	13,32	28,89	30,41	33,30	25,39	25,58
Standardabweichung	4,37	4,59	3,45	4,43	9,64	8,83	9,44	9,58	10,45
Glu Anzahl	323	182	97	44	1809	817	255	737	2132
niedrigster Wert	3,75	3,79	6,86	3,75	3,01	5,13	8,05	3,01	3,01
1. Quantil	10,69	10,56	11,60	7,92	19,60	20,90	23,23	16,21	16,10
Median	13,24	13,33	13,63	12,20	26,60	28,16	30,07	23,35	24,01
2. Quantil	15,39	15,70	15,11	14,73	33,49	34,14	35,81	30,09	32,11
höchster Wert	32,47	32,47	28,29	18,95	61,66	61,66	54,01	55,73	61,66
Durchschnittswert	13,15	13,22	13,75	11,50	26,75	28,16	30,12	24,02	24,69
Standardabweichung	3,98	4,22	3,33	3,93	9,66	9,06	8,98	9,84	10,27
Gly Anzahl	1885	1234	418	233	2501	999	301	1201	4386
niedrigster Wert	-7,39	-7,39	-4,76	-5,08	1,76	2,75	2,42	1,76	-7,39
1. Quantil	-0,18	-0,21	-0,22	0,09	11,82	12,32	15,52	10,64	1,68
Median	1,28	1,35	1,04	1,26	17,20	17,58	20,66	15,65	9,08
2. Quantil	2,69	2,67	2,63	2,81	22,88	23,44	24,76	21,53	18,62
höchster Wert	25,05	21,43	25,05	6,97	40,46	38,01	40,46	38,58	40,46
Durchschnittswert	1,35	1,34	1,41	1,33	17,53	17,97	20,19	16,49	10,57
Standardabweichung	2,57	2,50	2,97	2,18	7,10	7,00	6,46	7,13	9,78
His Anzahl	254	162	59	33	635	288	103	244	889
niedrigster Wert	2,94	5,86	5,29	2,94	3,52	6,06	5,66	3,52	2,94
1. Quantil	8,54	8,76	8,42	7,08	17,82	20,25	24,83	14,10	11,73
Median	10,55	10,55	10,73	9,52	24,95	26,95	29,76	20,31	19,60
2. Quantil	12,24	12,26	12,33	11,49	31,98	33,58	34,51	26,70	29,50
höchster Wert	21,87	21,87	18,17	17,81	52,52	50,12	52,52	46,33	52,52
Durchschnittswert	10,61	10,85	10,71	9,23	25,20	26,97	29,87	21,13	21,03
Standardabweichung	2,78	2,61	2,83	3,19	9,38	8,54	8,68	9,06	10,42
Ile Anzahl	2138	1723	215	200	1742	824	379	539	3880
niedrigster Wert	4,20	4,20	8,34	4,33	2,57	2,57	5,92	3,66	2,57
1. Quantil	11,92	11,95	12,51	10,71	12,39	12,28	14,82	11,02	12,07
Median	14,51	14,65	14,74	13,60	16,37	16,20	18,81	14,91	15,22
2. Quantil	17,53	17,64	18,47	16,12	20,38	19,99	21,57	19,17	18,83
höchster Wert	29,43	29,43	27,20	27,39	34,75	34,75	32,85	29,58	34,75
Durchschnittswert	14,86	14,93	15,49	13,54	16,32	16,19	18,39	15,06	15,52
Standardabweichung	3,91	3,86	4,04	3,96	5,34	5,20	4,86	5,44	4,66

Leu Anzahl	3209	2486	478	245	2786	1419	440	927	5995
niedrigster Wert	4,11	5,15	7,21	4,11	2,35	4,77	5,49	2,35	2,35
1. Quantil	13,81	13,86	13,94	12,29	14,24	14,69	17,15	12,26	13,94
Median	16,99	17,07	16,80	16,24	19,31	19,43	21,21	17,64	17,81
2. Quantil	20,29	20,34	19,94	20,05	23,50	23,76	24,91	22,43	21,73
höchster Wert	36,69	30,68	33,94	36,69	40,16	37,58	35,25	40,16	40,16
Durchschnittswert	17,21	17,30	17,20	16,33	18,98	19,31	21,02	17,51	18,03
Standardabweichung	4,47	4,41	4,33	5,25	6,38	6,09	5,86	6,71	5,51
Lys Anzahl	319	186	98	35	1873	816	322	735	2192
niedrigster Wert	5,65	5,65	9,67	7,15	3,26	3,26	10,95	3,42	3,26
1. Quantil	18,92	17,72	20,93	13,22	23,57	26,05	30,11	19,65	22,31
Median	22,71	21,57	26,17	20,31	31,92	33,16	37,01	27,19	30,14
2. Quantil	27,19	26,30	28,44	22,93	40,05	40,91	43,29	36,19	38,86
höchster Wert	40,22	40,22	39,47	35,75	65,35	64,15	59,86	65,35	65,35
Durchschnittswert	22,98	22,46	25,17	19,63	32,15	33,42	36,61	28,78	30,81
Standardabweichung	5,97	5,78	5,30	6,71	10,98	10,11	9,40	11,56	10,90
Met Anzahl	667	516	84	67	646	333	97	216	1313
niedrigster Wert	3,63	4,25	6,52	3,63	3,46	5,16	11,62	3,46	3,46
1. Quantil	10,16	10,13	10,27	10,17	16,66	18,04	20,60	14,06	11,67
Median	12,39	12,49	12,14	12,22	21,85	22,90	24,00	19,47	15,27
2. Quantil	14,76	14,76	15,02	14,17	26,28	26,29	29,20	24,59	21,86
höchster Wert	26,36	23,67	22,86	26,36	37,81	36,06	37,37	37,81	37,81
Durchschnittswert	12,56	12,58	12,55	12,43	21,61	22,09	24,66	19,50	17,01
Standardabweichung	3,42	3,40	3,02	4,03	6,71	6,17	6,15	7,11	6,97
Phe Anzahl	1689	1299	237	153	1430	679	259	492	3119
niedrigster Wert	2,10	3,37	4,37	2,10	2,11	3,16	4,08	2,11	2,10
1. Quantil	9,24	9,26	9,35	8,76	10,58	10,87	12,86	9,13	9,69
Median	11,38	11,45	11,42	11,11	14,46	14,46	15,99	13,07	12,37
2. Quantil	13,58	13,57	13,47	13,70	18,01	17,92	19,38	17,18	15,58
höchster Wert	24,46	24,46	20,72	21,57	31,93	28,50	31,93	29,46	31,93
Durchschnittswert	11,61	11,66	11,60	11,19	14,40	14,53	16,16	13,30	12,89
Standardabweichung	3,26	3,24	3,15	3,62	5,16	4,90	4,88	5,38	4,46
Pro Anzahl	568	400	57	111	1663	625	197	841	2231
niedrigster Wert	0,41	0,41	3,54	1,75	2,59	4,47	4,57	2,59	0,41
1. Quantil	6,66	6,91	5,85	6,08	15,69	17,13	17,56	14,13	10,49
Median	8,63	8,96	7,52	7,83	22,12	23,67	23,81	20,25	17,79
2. Quantil	10,32	10,45	9,67	9,60	28,94	29,87	30,61	27,95	26,62
höchster Wert	37,30	37,30	15,26	13,33	53,31	51,12	53,31	52,85	53,31
Durchschnittswert	8,75	9,08	8,10	7,93	22,66	23,93	24,21	21,36	19,12
Standardabweichung	3,29	3,49	2,70	2,62	8,81	8,31	8,84	8,97	9,86
Ser Anzahl	1225	824	247	154	2105	840	349	916	3330
niedrigster Wert	-2,15	-2,15	-1,92	-1,48	1,55	3,91	5,38	1,55	-2,15
1. Quantil	2,11	2,27	2,17	1,54	13,73	14,89	16,72	11,89	4,42
Median	3,52	3,57	3,44	2,80	19,04	19,70	22,79	16,90	12,97
2. Quantil	4,91	4,86	5,15	4,44	24,49	24,74	27,36	22,55	21,24
höchster Wert	35,30	35,30	28,59	12,17	45,35	36,58	37,17	45,35	45,35
Durchschnittswert	3,77	3,79	4,04	3,22	19,25	19,96	22,20	17,48	13,56
Standardabweichung	2,83	2,79	3,19	2,38	7,10	6,59	6,78	7,20	9,52

Thr Anzahl	1130	790	207	133	1947	819	345	783	3077
niedrigster Wert	1,14	3,02	4,42	1,14	3,21	4,74	3,42	3,21	1,14
1. Quantil	8,09	8,18	8,52	6,46	15,42	15,71	18,92	13,75	10,32
Median	10,01	10,06	10,36	8,98	20,67	20,84	23,19	19,01	15,17
2. Quantil	11,98	11,94	12,27	11,61	25,65	25,77	27,45	24,48	22,76
höchster Wert	34,01	34,01	21,78	19,04	48,34	48,34	38,55	40,15	48,34
Durchschnittswert	10,18	10,26	10,56	9,12	20,64	20,98	23,08	19,21	16,80
Standardabweichung	3,09	3,00	2,91	3,62	7,06	6,66	6,60	7,34	7,78
Trp Anzahl	521	382	94	45	513	247	99	167	1034
niedrigster Wert	3,21	6,02	3,21	4,57	3,16	4,95	6,47	3,16	3,16
1. Quantil	12,91	13,04	13,64	9,51	12,59	12,81	16,04	11,05	12,72
Median	15,58	15,76	16,12	12,92	16,80	16,70	19,76	15,39	16,17
2. Quantil	18,43	18,59	18,17	16,13	21,01	21,09	22,38	19,04	19,76
höchster Wert	29,90	29,90	27,25	23,15	32,43	31,15	32,43	31,43	32,43
Durchschnittswert	15,69	15,90	16,17	12,87	17,09	17,15	19,88	15,35	16,38
Standardabweichung	4,08	3,91	4,15	4,40	5,97	5,74	5,52	5,94	5,15
Tyr Anzahl	846	514	237	95	1159	521	261	377	2005
niedrigster Wert	4,10	4,46	5,75	4,10	2,63	4,62	5,29	2,63	2,63
1. Quantil	9,93	10,13	9,99	8,08	14,72	15,03	18,21	11,70	11,51
Median	12,25	12,52	12,30	10,83	19,37	19,67	21,67	17,18	15,33
2. Quantil	14,62	14,93	14,42	13,02	23,53	23,66	25,51	21,68	20,69
höchster Wert	28,05	28,05	25,71	24,72	40,13	40,13	39,05	39,56	40,13
Durchschnittswert	12,51	12,79	12,56	10,88	19,13	19,33	21,88	16,97	16,34
Standardabweichung	3,59	3,54	3,49	3,72	6,19	5,76	5,54	6,38	6,19
Val Anzahl	2217	1734	301	182	2060	920	499	641	4277
niedrigster Wert	2,55	4,45	3,46	2,55	2,46	4,03	3,17	2,46	2,46
1. Quantil	10,28	10,38	9,90	9,39	12,05	12,08	14,92	10,20	10,81
Median	12,78	12,84	12,78	12,10	15,99	15,91	18,04	13,97	14,10
2. Quantil	15,66	15,71	15,91	14,88	19,82	19,60	21,17	18,05	17,57
höchster Wert	29,36	27,44	29,36	25,43	29,67	29,30	29,67	28,59	29,67
Durchschnittswert	13,11	13,19	13,22	12,18	15,85	15,84	18,01	14,18	14,43
Standardabweichung	3,77	3,61	4,41	4,08	5,28	5,09	4,64	5,41	4,76
alle Residuen Anzahl	21308	15385	3653	2270	32081	14348	5211	12522	553389
niedrigster Wert	-7,39	-7,39	-4,76	-5,08	0,89	1,07	2,42	0,89	-7,39
1. Quantil	6,90	7,24	6,45	5,37	14,39	15,06	17,17	12,68	10,77
Median	11,50	11,68	11,67	9,87	20,02	20,60	22,37	18,24	15,99
2. Quantil	15,53	15,67	15,80	14,11	26,52	26,95	29,09	24,74	22,55
höchster Wert	43,75	43,75	39,47	36,69	65,35	64,15	61,66	65,35	65,35
Durchschnittswert	11,37	11,57	11,43	9,97	21,22	21,81	23,75	19,49	17,29
Standardabweichung	6,29	6,18	6,80	6,04	9,33	9,18	9,23	9,19	9,56

Tabelle 14: Korrelationsmatrix für innerhalb der Membran stehende Aminosäuren.

Zur besseren Übersicht wurden die Werte eingefärbt. Der höchste Wert der Matrix trägt die Farbe Grün, der niedrigste Wert die Farbe Rot. Ambivalente Werte sind gelb eingefärbt.

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	1,00	-0,07	0,05	0,12	0,86	-0,01	0,12	0,28	0,31	-0,01	-0,09	-0,16	0,15	0,24	0,68	0,74	0,43	-0,04	0,16	0,12
Arg	-0,07	1,00	0,67	0,67	-0,01	0,82	0,59	-0,15	0,28	0,81	0,95	0,70	0,53	0,42	0,10	-0,12	0,25	0,88	0,53	0,62
Asn	0,05	0,67	1,00	0,97	0,17	0,92	0,98	-0,12	0,67	0,95	0,80	0,19	0,93	0,82	0,39	-0,07	0,62	0,89	0,91	0,97
Asp	0,12	0,67	0,97	1,00	0,28	0,91	0,98	-0,11	0,75	0,93	0,76	0,18	0,94	0,87	0,51	-0,04	0,71	0,87	0,93	0,97
Cys	0,86	-0,01	0,17	0,28	1,00	0,08	0,25	0,09	0,52	0,09	-0,04	-0,14	0,31	0,45	0,88	0,42	0,67	0,05	0,34	0,29
Gln	-0,01	0,82	0,92	0,91	0,08	1,00	0,86	-0,14	0,53	0,96	0,90	0,35	0,82	0,71	0,27	-0,10	0,49	0,94	0,81	0,86
Glu	0,12	0,59	0,98	0,98	0,25	0,86	1,00	-0,11	0,76	0,91	0,72	0,13	0,95	0,88	0,51	-0,04	0,71	0,84	0,94	0,96
Gly	0,28	-0,15	-0,12	-0,11	0,09	-0,14	-0,11	1,00	-0,08	-0,14	-0,16	-0,18	-0,11	-0,10	0,00	0,72	-0,06	-0,14	-0,11	-0,12
His	0,31	0,28	0,67	0,75	0,52	0,53	0,76	-0,08	1,00	0,64	0,36	-0,06	0,85	0,95	0,83	0,04	0,97	0,48	0,89	0,85
Ile	-0,01	0,81	0,95	0,93	0,09	0,96	0,91	-0,14	0,64	1,00	0,90	0,35	0,89	0,79	0,34	-0,11	0,58	0,96	0,88	0,92
Leu	-0,09	0,95	0,80	0,76	-0,04	0,90	0,72	-0,16	0,36	0,90	1,00	0,61	0,66	0,53	0,11	-0,14	0,32	0,96	0,64	0,72
Lys	-0,16	0,70	0,19	0,18	-0,14	0,35	0,13	-0,18	-0,06	0,35	0,61	1,00	0,10	0,02	-0,11	-0,17	-0,06	0,46	0,09	0,17
Met	0,15	0,53	0,93	0,94	0,31	0,82	0,95	-0,11	0,85	0,89	0,66	0,10	1,00	0,96	0,61	-0,04	0,83	0,76	0,99	0,97
Phe	0,24	0,42	0,82	0,87	0,45	0,71	0,88	-0,10	0,95	0,79	0,53	0,02	0,96	1,00	0,75	0,00	0,94	0,65	0,98	0,94
Pro	0,68	0,10	0,39	0,51	0,88	0,27	0,51	0,00	0,83	0,34	0,11	-0,11	0,61	0,75	1,00	0,26	0,92	0,22	0,65	0,60
Ser	0,74	-0,12	-0,07	-0,04	0,42	-0,10	-0,04	0,72	0,04	-0,11	-0,14	-0,17	-0,04	0,00	0,26	1,00	0,09	-0,11	-0,04	-0,06
Thr	0,43	0,25	0,62	0,71	0,67	0,49	0,71	-0,06	0,97	0,58	0,32	-0,06	0,83	0,94	0,92	0,09	1,00	0,44	0,86	0,81
Trp	-0,04	0,88	0,89	0,87	0,05	0,94	0,84	-0,14	0,48	0,96	0,96	0,46	0,76	0,65	0,22	-0,11	0,44	1,00	0,76	0,83
Tyr	0,16	0,53	0,91	0,93	0,34	0,81	0,94	-0,11	0,89	0,88	0,64	0,09	0,99	0,98	0,65	-0,04	0,86	0,76	1,00	0,98
Val	0,12	0,62	0,97	0,97	0,29	0,86	0,96	-0,12	0,85	0,92	0,72	0,17	0,97	0,94	0,60	-0,06	0,81	0,83	0,98	1,00

Tabelle 15: Korrelationsmatrix für außerhalb der Membran stehende Aminosäuren.

Zur besseren Übersicht wurden die Werte eingefärbt. Der höchste Wert der Matrix trägt die Farbe Grün, der niedrigste Wert die Farbe Rot. Ambivalente Werte sind gelb eingefärbt.

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	1,00	0,27	0,63	0,77	0,54	0,33	0,54	0,95	0,60	0,99	0,93	0,20	0,76	0,94	0,80	0,92	0,86	0,96	0,92	0,99
Arg	0,27	1,00	0,84	0,74	-0,18	0,95	0,90	0,37	0,82	0,24	0,46	0,95	0,64	0,09	0,72	0,49	0,58	0,30	0,46	0,20
Asn	0,63	0,84	1,00	0,96	0,05	0,89	0,97	0,70	0,93	0,60	0,79	0,79	0,89	0,45	0,94	0,81	0,88	0,65	0,79	0,56
Asp	0,77	0,74	0,96	1,00	0,15	0,80	0,93	0,82	0,92	0,74	0,90	0,67	0,95	0,58	0,97	0,90	0,95	0,78	0,88	0,70
Cys	0,54	-0,18	0,05	0,15	1,00	-0,13	-0,02	0,60	0,07	0,53	0,33	-0,24	0,13	0,72	0,26	0,42	0,25	0,50	0,29	0,56
Gln	0,33	0,95	0,89	0,80	-0,13	1,00	0,94	0,42	0,88	0,30	0,51	0,92	0,68	0,17	0,77	0,56	0,64	0,36	0,51	0,27
Glu	0,54	0,90	0,97	0,93	-0,02	0,94	1,00	0,62	0,93	0,51	0,72	0,84	0,84	0,36	0,91	0,74	0,82	0,57	0,71	0,47
Gly	0,95	0,37	0,70	0,82	0,60	0,42	0,62	1,00	0,66	0,93	0,94	0,26	0,83	0,89	0,85	0,96	0,90	0,93	0,91	0,92
His	0,60	0,82	0,93	0,92	0,07	0,88	0,93	0,66	1,00	0,57	0,73	0,78	0,84	0,43	0,91	0,76	0,82	0,61	0,71	0,54
Ile	0,99	0,24	0,60	0,74	0,53	0,30	0,51	0,93	0,57	1,00	0,91	0,18	0,73	0,94	0,77	0,90	0,83	0,96	0,89	0,99
Leu	0,93	0,46	0,79	0,90	0,33	0,51	0,72	0,94	0,73	0,91	1,00	0,38	0,93	0,77	0,88	0,96	0,97	0,91	0,98	0,89
Lys	0,20	0,95	0,79	0,67	-0,24	0,92	0,84	0,26	0,78	0,18	0,38	1,00	0,54	0,03	0,63	0,40	0,49	0,23	0,37	0,15
Met	0,76	0,64	0,89	0,95	0,13	0,68	0,84	0,83	0,84	0,73	0,93	0,54	1,00	0,55	0,92	0,91	0,96	0,78	0,92	0,69
Phe	0,94	0,09	0,45	0,58	0,72	0,17	0,36	0,89	0,43	0,94	0,77	0,03	0,55	1,00	0,65	0,80	0,68	0,91	0,75	0,96
Pro	0,80	0,72	0,94	0,97	0,26	0,77	0,91	0,85	0,91	0,77	0,88	0,63	0,92	0,65	1,00	0,92	0,94	0,80	0,87	0,74
Ser	0,92	0,49	0,81	0,90	0,42	0,56	0,74	0,96	0,76	0,90	0,96	0,40	0,91	0,80	0,92	1,00	0,96	0,92	0,95	0,88
Thr	0,86	0,58	0,88	0,95	0,25	0,64	0,82	0,90	0,82	0,83	0,97	0,49	0,96	0,68	0,94	0,96	1,00	0,85	0,96	0,80
Trp	0,96	0,30	0,65	0,78	0,50	0,36	0,57	0,93	0,61	0,96	0,91	0,23	0,78	0,91	0,80	0,92	0,85	1,00	0,89	0,95
Tyr	0,92	0,46	0,79	0,88	0,29	0,51	0,71	0,91	0,71	0,89	0,98	0,37	0,92	0,75	0,87	0,95	0,96	0,89	1,00	0,87
Val	0,99	0,20	0,56	0,70	0,56	0,27	0,47	0,92	0,54	0,99	0,89	0,15	0,69	0,96	0,74	0,88	0,80	0,95	0,87	1,00

Literaturverzeichnis

- 1 Diehn, M.; Bhattacharya, R.; Botstein, D.; Brown, P.: Genome-scale identification of membrane associated human mRNAs. In: PLoS Genet. - United States : Public Library of Science., 2 (1), S. e11 (2006)
- 2 Rader, A. et al.: Identification of core amino acids stabilizing rhodopsin. In: Proc Natl Acad Sci U S A. - United States : National Academy of Sciences. - 101 (19), S. 7246-7251 (2004)
- 3 Du, K.; Sharma, M.; Lukacs, G.: The deltaf508 cystic fibrosis mutation. In: Nat Struct Mol Biol. - United States : Nature Pub. Group. - 12 (1), S. 17-25 (2005)
- 4 Tanzi, R.; Kovacs, D.; Kim, T.; Moir, R.; Guenette, S.; Wasco, W.: The gene defects responsible for familial Alzheimer's disease. In: Neurobiol Dis. - United States : Academic Press. - 3 (3), S. 159-168 (1996)
- 5 Soussi, T.; Bérout, C.: Assessing TP53 status in human tumours to evaluate clinical outcome. In: Nat Rev Cancer. - England : Nature Pub. Group. - 1 (3), S. 233-240 (2001)
- 6 Müller-Esterl, Werner. Biochemie: Eine Einführung für Mediziner und Naturwissenschaftler. - 1. Auflage. Heidelberg : Spektrum Akademischer Verlag, 2009.
- 7 URL:
http://www.dsimb.inserm.fr/~debrevern/VENN_DIAGRAM/aa_venn_diagram.png, 11.06.2010
- 8 Merkel, Rainer; Waack, Stephan: Bioinformatik interaktiv. - 2. Auflage. Weinheim : Wiley-VCH Verlag GmbH & Co. KGaA, 2009
- 9 Löffler, Georg: Basiswissen Biochemie mit Pathobiochemie. - 6. Auflage. Heidelberg : Springer-Verlag GmbH, 2004
- 10 Dose, Klaus: Biochemie. - 5. Auflage. Heidelberg : Springer-Verlag GmbH, 1994
- 11 Deckwer, Rolf-Dieter: Römpp-Lexikon: Biotechnologie und Gentechnik. - Stuttgart : Georg Thieme Verlag, 1999
- 12 Gibas, Cynthia: Einführung in die Praktische Bioinformatik. - Köln : O'Reilly, 2002
- 13 URL: <http://www.u-helmich.de/che/09/05-molek/Folien/Fo12-06.jpg>, 15.06.2010
- 14 URL: http://www.baertierchen.de/f_glyc.jpg, 22.07.2010
- 15 Folienserie Super Quintet III-57.
- 16 URL: <http://www.enzim.hu/~szia/folding/ea02/golfg.gif>, 15.06.2010
- 17 Levinthal, Cyrus: Are there pathways for protein folding. In: J. Chim. Phys 65 (1), S. 44 (1968)

- 18 Nienhaus, Gerd Ulrich: Physik der Proteine. In: Physik Journal. - Weinheim : WILEY-VCH Verlag GmbH & Co. KG. - 3 (4), S. 37-43 (2004)
- 19 Bryngelson, Joseph D., et al. Funnels, Pathways and the Energy Landscape of Protein Folding: A. Proteins: In: Structure, Function, and Bioinformatics. - United States : Wiley-Liss. - 21 (3), S. 167-195 (2004)
- 20 Ghaemmaghani, S.; Oas, T.: Quantitative proteinstability measurement in vivo. In: Nat Struct Biol. - United States : Nature Pub. Co. 8 (10), S. 879-883 (2001)
- 21 Graña, O.; Baker, D.; MacCallum, R.; Meiler, J.; Punta, M.; Rost, B.; Tress, M.; Valencia, A.: Casp6 assessment of contact prediction. In: Proteins. - United States : Wiley-Liss. - 61 (S7), S. 214-224 (2005)
- 22 Dressel, Frank. Sequenz, Energie, Struktur - Untersuchungen zur Beziehung zwischen Primär- und Tertiärstruktur in globulären und Membran-Proteinen. Dissertation. Mathematisch-Naturwissenschaftlichen Fakultät der Technischen Universität Dresden : s.n., 2008.
- 23 Fraternali, Franca; Cavallo, Luigi: POPS. URL: <http://mathbio.nimr.mrc.ac.uk/wiki/POPS>, 20.08.2009.
- 24 Banavar, J. et al.: Structural motifs of biomolecules. In: Proc Natl Acad Sci U S A. - United States : National Academy of Sciences. - 104 (44), S. 17283-17286 (2007)
- 25 Institute of Enzymology, Budapest, Hungary: Protein Data Bank of Transmembrane Proteins. URL: <http://pdbtm.enzim.hu>, 02.03.2007
- 26 European Bioinformatics Institute. PDBsum. URL: <http://www.ebi.ac.uk/pdbsum/>, 24.08.2010
- 27 Unwin, N.; Henderson, R.: The structure of proteins in biological membranes. In: Scientific American. - United States : Scientific American. - 250 (2), S. 78-94 (1984)

Danksagung

An erster Stelle gebührt mein Dank Prof. Dr. Dirk Labudde, der diese Arbeit ermöglichte und mich bei Bedarf mit wertvollen Ratschlägen unterstützte. Weiterhin bedanke ich mich bei meinem Zweitbetreuer Dipl. Inf. (FH) Daniel Stockmann für die Unterstützung in kniffligen Situationen, vor allem im programmiertechnischen Bereich. Bei meinen Kommilitonen Florian Heinke und Eric Frenzel möchte ich mich für die zahlreichen, fruchtbaren Diskussionen, im Hinblick auf die Energieprofilthematik, bedanken.

Besonders danken möchte ich auch meiner Freundin, Marie Karsunke, die mir ihren Laptop zur Verfügung stellte, als mein PC den Geist aufgab.

Höchste Wertschätzung gebührt meiner Familie, die mir in schwierigen Situationen Rückhalt gegeben hat, und mich eifrig im Kampf gegen Druckfehler und Stilblüten unterstützt hat.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Wüstenbrand, 22.08.2010

Riccardo Brumm